EXPLICIT AND INCIDENTAL INSTRUCTION

AND LEARNER AWARENESS

A THESIS SUBMITTED TO THE GRADUATE DIVIDSION OF THE
UNIVERSITY OF HAWAIʻI IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

IN

ENGLISH AS A SECOND LANGUAGE

MAY 1999

By
Eric K. Hauser

Thesis Committee:

Richard Schmidt, Chairperson
Robert Bley-Vroman
James Dean Brown

Table 2.1: Experimental Effects of Instruction Studies

Study	Type (classroom or laboratory; natural or artificial language)	Contrast	Results	Measure of Learner-Internal Processes? (learner awareness, behavior)
Doughty (1991)	laboratory; natural language (English)	meaning-focused vs. form-focused vs. exposure only on acquisition of relative clauses	improvement in relativization by both instruction groups; improved compre-hension by meaning group	no
Kepner (1991)	classroom; natural language (Spanish)	message-related vs. corrective feedback on learners' writing	improved surface accuracy and ideational quality only for learners receiving message-related feedback	no
Mondria & Wit-de Boer (1991)	classroom; natural language (French)	inference-conducive context vs. lack of such context on vocabulary acquisition	improved inference but not retention in inference-conducive context	no
Rose Chang & Smith (1991)	classroom; natural language (Spanish)	individual learning vs. cooperative learning on learners' achievement	no effect on overall achievement	yes (audio recordings of dyads showing how they approached the task; nothing for monads)
White (1991)	classroom; natural language (English); children	explicit (including negative feedback) vs. positive evidence alone on adverb placement	short-term improvement with explicit instruction, effects not lasting	no

Table 2.1 (continued): Experimental Effects of Instruction Studies

Study	Type (classroom or laboratory; natural or artificial language)	Contrast	Results	Measure of Learner-Internal Processes? (learner awareness, behavior)
White, et. al (1991)	classroom; natural language (English); children	explicit instruction + feedback vs. positive evidence without instruction on question formation	improvement with explicit instruction, still maintained after five weeks	no
Carroll & Swain (1993)	laboratory; natural language (English)	explicit vs. implicit corrective feedback vs. no feedback on acquisition of dative alternation	improvement for all feedback groups, especially for explicit feedback	no
DeKeyser (1993)	classroom; natural language (French)	explicit corrective feedback vs. no corrective feedback during oral production on proficiency	no overall effect; error correction beneficial for learners with high previous achievement, low anxiety, and low extrinsic motivation	no (however, questionnaire used to assess extrinsic motivation and anxiety)
N. Ellis (1993)	laboratory; natural language (Welsh)	exposure to rules vs. instances vs. rules + instances on acquisition of orthographic rules	limited learning with instances; better learn- with rules; best learn- with rules and instances	no
Fotos (1993)	classroom; natural language (English)	effect of grammar lesson vs. grammar task vs. communicative task on noticing	grammar lesson and grammar task result in more noticing	yes (instruction to underline "special use of English" in texts to assess noticing)

Table 2.1 (continued): Experimental Effects of Instruction Studies

Study	Type (classroom or laboratory; natural or artificial language)	Contrast	Results	Measure of Learner-Internal Processes? (learner awareness, behavior)
VanPatten & Cadierno (1993)	classroom; natural language (Spanish)	explicit vs. input-processing instruction on acquisition of clitic direct object pronouns	improved production with explicit instruction; improved production and comprehension with input-processing instruction	no
R. Ellis, et. al (1994)	classroom; natural language (English)	premodified vs. interactionally-modified vs. unmodified input on vocabulary acquisition	modified input pro-acquisition; inter-actionally-modified better than premodified	no
Fotos (1994)	classroom; natural language (English)	grammar conscious-ness raising tasks vs. communicative tasks and vs. teacher-fronted instruction on word order	improved accuracy with grammar tasks compar-able with teacher-fronted instruction; quantity of interaction with grammar tasks comparable with communicative tasks	no
Master (1994)	classroom; natural language (English)	10-week systematic instruction vs. no instruc-tion on acquisition of articles	improvement on article use for learners receiving instruction; no improvement without instruction	no

Table 2.1 (continued): Experimental Effects of Instruction Studies

Study	Type (classroom or laboratory; natural or artificial language)	Contrast	Results	Measure of Learner-Internal Processes? (learner awareness, behavior)
Alanen (1995)	laboratory; semi-artificial (Finnish-based)	rules vs. enhancement vs. rules + enhancement vs. exposure only on acquisition of locative suffixes and consonant alternation	treatments facilitated noticing and acquisition	yes (think aloud protocols; questionnaire asking for rule statements and learners' perception of enhancement)
Cadierno (1995)	classroom; natural language (Spanish)	processing instruction vs. traditional instruction vs. no instruction on acquisition of past tense verb morphology	improved production with traditional instruction; improved production and comprehension with input-processing instruction	no
DeKeyser (1995)	laboratory; artificial language	explicit-deductive vs. implicit-inductive instruction on acquisition of categorical and prototypical morphological rules	categorical rules learned better with explicit instruction; prototypical rules learned better with implicit instruction	yes (retrospective interview investigating how learners approached the task)
R. Ellis (1995)	classroom; natural language (English)	premodified input vs. interactionally-modified input on vocabulary acquisition	acquisition with both types of input, but better for interacionally-modified input	yes (uptake charts asking only learners receiving interactionally modified input about new vocabulary)

Table 2.1 (continued): Experimental Effects of Instruction Studies

Study	Type (classroom or laboratory; natural or artificial language)	Contrast	Results	Measure of Learner-Internal Processes? (learner awareness, behavior)
Hulstijn, et. al (1996)	laboratory; natural language (French)	marginal gloss vs. dictionary use vs. no additional information and single vs. multiple occurrence in incidental vocabulary acquisition	greater retention with gloss and dictionary (at least when consulted), especially for multiple-occurrence vocabulary	yes (word recognition test to assess whether learners were aware of words that had appeared in the story; indication by dictionary users of looked-up words)
Robinson (1995,1996a, 1996b, 1997a)	laboratory; natural language (English)	implicit vs. incidental vs. rule-search vs. instructed conditions on easy and hard rules	superior performance by learners in instructed condition	yes (retrospective interview investigating how learners had approached the task; rule verbalization)
Trahey (1996)	classroom; natural language (English); children	input flood vs. explicit instruction and no instruction on adverb placement*	input flood more effective for long-term retention; however, ungrammatical placement still accepted	no
Winitz (1996)	classroom; natural language (Spanish)	implicit instruction vs. explicit instruction on grammaticality judgements	more accurate judgements with implicit instruction	no
de Graaff (1997)	laboratory; artificial language (Spanish-based)	explicit vs. implicit instruction on morpho-syntax	explicit instruction facilitates acquisition	no (results of debriefing interview not mentioned)

Table 2.1 (continued): Experimental Effects of Instruction Studies

Study	Type (classroom or laboratory; natural or artificial language)	Contrast	Results	Measure of Learner-Internal Processes? (learner awareness, behavior)
DeKeyser (1997)	laboratory; artificial language	production practice vs. comprehension practice (both following explicit instruction) on morphosyntax	production practice facilitates production; comprehension practice facilitates comprehension	yes (metalinguistic test to determine whether criterion level had been reached)
Robinson (1997b)	laboratory; semi-artificial (English-based)	implicit vs. incidental vs. enhanced vs. instructed conditions on argument structure of novel verbs	superior performance in enhanced and instructed conditions	yes (retrospective interview investigating how learners approached the task; rule verbalization)
Tinkham (1997)	laboratory; artificial vocabulary	semantic-clustering vs. thematic-clustering on vocabulary acquisition	superiority of thematic-clustering	yes (debriefing interview asking for learners' perception of ease and difficulty of learning)
Yang & Givon (1997)	laboratory; artificial language	grammatical input vs. simplified pidgin input on acquisition of vocabulary and grammar	pidgin input does not facilitate vocabulary acquisition, but hinders grammar acquisition	no (however questionnaire used to assess learners' motivation)
Zimmerman, C. M. (1997)	classroom; natural language (English)	self-selected reading with interactive vocabulary instruction vs. without on vocabulary acquisition	beneficial effects for interactive vocabulary instruction	yes (pre and post questionnaires used to assess learners perceptions of vocabulary learning strategies)

Table 2.1 (continued): Experimental Effects of Instruction Studies

Study	Type (classroom or laboratory; natural or artificial language)	Contrast	Results	Measure of Learner-Internal Processes? (learner awareness, behavior)
Izumi & Lakshmanan (1998)	classroom (specially-created for the research); natural language (English)	provision of negative evidence vs. no negative evidence on acquisition of passive	improvement with provision of negative evidence	no
Joe (1998)	classroom (specially-created for the research); natural language (English)	explicit training to generate recently encountered vocabulary during retelling vs. no training on vocabulary acquisition	generation facilitates acquisition; training does not result in more generation	yes (self-reports indicating learners' familiarity with vocabulary)
Leow (1998)	classroom; natural language (Spanish)	learner-centered vs. teacher-centered exposure and single vs. multiple exposure acquisition of stem-changing verbs	benefits for learner-centered and multiple exposure	yes (think-aloud protocols to assess noticing and activation of prior knowledge)
Long, et. al (1998)	laboratory; natural languages (Japanese & Spanish)	preemptive positive input (models) vs. implicit negative feedback (recasts) on acquisition of (untrue) word order (Japanese) and fronting of direct objects	some evidence that adults can learn from recasts and that recasts are more effective	no

*The explicit instruction and no instruction groups are from White (1991).

Under the heading of *Type,* the table lists information on whether the study was conducted in a laboratory or a classroom and whether it involved a natural or artificial language. Laboratory studies allow for a greater degree of control than do classroom studies, but may also be less generalizable to classroom contexts. The same is true for artificial versus natural languages. The table also indicates the instruction contrast that the study investigated. Only experimental studies which contrasted two or more kinds of instruction are included in this table. The basic results are given next. The original studies should be consulted for more detailed information about the contrast and results. Finally, under the heading of *Measure of Learner-Internal Processes,* information is given about whether there was an attempt in the study to measure learner awareness, noticing, or behavior. An example of an assessment of learner behavior is DeKeyser (1995), in which learners were interviewed to see to what extent they followed task directions.

As Table 2.1 shows, researchers have taken a variety of approaches in studying the effects of instruction. Of the 32 studies, 13 have taken place in laboratories and nineteen in classrooms, with one trend being a shift toward laboratories as the decade progresses. Twenty-five of the studies have used natural languages, with 14 studies using English, and seven studies have used artificial languages. Here also, there is a second trend, a shift toward the use of artificial languages as the decade progresses. Researchers have tended to investigate the effects of different kinds of instruction on morphosyntax, but writing, orthography, and vocabulary acquisition have also been investigated. Results are mixed, but tend to indicate beneficial effects for some form of instruction which focuses on both form and meaning (that is, focus on form, Doughty & Williams, 1998).

Some of these studies are well-designed and very informative about the effects of different kinds of instruction (e.g., Doughty, 1991; VanPatten & Cadierno, 1993; Cadierno, 1995; Winitz, 1996; Long et al., 1998), but because they do not include any measures of learner awareness or learner behavior in response to the instruction, they do not illuminate any learner-internal processes. These studies are critiqued in this section in order to illustrate the questions that they leave open about learner awareness and behavior. This is, of course, rather unfair to these researchers since they were not actually trying to investigate these learner-internal processes. However, it is done anyway in order to speculate on what these studies may have revealed had awareness and behavior been investigated.

A third trend that can be seen in the table, the trend which is most relevant to this paper, is a growing number of studies, as the decade progresses, that do investigate learner-internal processes. Some of these studies are also critiqued for how well they illuminate the role of learner-internal processes in SLA.

No measure of learner-internal processes Doughty (1991) conducted a tightly controlled study of the learning of English relativization by learners receiving three different types of computerized instruction. A meaning-oriented group not only viewed sentences containing object of a preposition relative clauses but was also given information to help them understand the meaning of the sentences and presented with highlighting and capitalization features to draw their attention to structure. A rule-oriented group was presented with explicit rules and computer animation which deconstructed and reconstructed object of a preposition relative clauses so that learners could see how they are formed. A control group was merely exposed to the sentences.

Doughty found that both the meaning-oriented group and the rule-oriented group showed more improvement that the control group on relativization ability, while only the meaning-oriented group showed improvement in comprehension of sentences containing relative clauses.

Doughty attributes the ability of the meaning-oriented group to improve not only in comprehension but also on the formal features of relativization to the techniques (highlighting and capitalization) used to make the features salient and draw learners attention. This is certainly a plausible explanation. However, without some means of discovering exactly what drew the learners' attention or what the learners were aware of having learned, it is impossible to know for sure how effective the techniques were in focusing learners' attention on form, or how aware of form learners were. In addition, Doughty found that learners in the meaning-oriented group and the rule-oriented group, as a result of receiving instruction on object of a preposition relative clauses, also improved on less-marked relative clauses, as well as, quite surprisingly, more-marked relative clauses. This seems to indicate that the learners learned more than they had actually been taught, begging "the question of how this apparent advantage for SLA occurs" (p. 464). It would seem that investigating what learners actually become aware of during the process of learning may illuminate the origins of this advantage.

VanPatten and Cadierno (1993) compared two groups of learners receiving instruction on the use of clitic direct object pronouns in Spanish. One group, a *traditional instruction* group, was taught explicit rules on the formation of sentences with clitic direct object pronouns and completed exercises focusing on the production of sentences containing the pronouns. The exercises themselves progressed from mechanical drills, to meaningful practice, to communicative practice. The other group, an *input processing* group, received explanations of how to interpret the meaning of sentences containing clitic direct object pronouns and practiced responding to the content of sentences containing the pronouns. Both groups received some kind of explicit instruction, but while the traditional instruction group received rules for and practiced production, the input processing group received help in processing and interpretation. Cadierno (1995) conducted a similar study on the acquisition of past tense verb morphology in Spanish.

Interestingly, VanPatten and Cadierno (1993) and Cadierno (1995) found that while learners in the traditional instruction group improved only in production, learners in the input processing group improved in both production and comprehension. They attribute this to input processing instruction helping learners' processing mechanisms convert input to intake which could then be integrated into learners' interlanguage systems. Traditional instruction, though, was claimed to result in learned linguistic knowledge which could only be used to monitor output in production. However, because VanPatten and Cadierno did not attempt to discover what learners became aware of during instruction, it is again impossible to know what they were focusing on during instruction. Were learners in the input processing group solely focusing on comprehension, or were they also trying to figure out rules for producing sentences with clitic direct object pronouns? Were all learners in the traditional instruction group focusing exclusively on producing formally accurate sentences, or were they also directing their attention toward meaning?

Of the studies listed in Table 2.0, Winitz (1996) is unusual in that it found benefits for what the author terms implicit instruction rather than for explicit instruction. In this study, participants were given a grammaticality judgement test during the last week of their first semester of Spanish. About half of these learners had been attending grammar/translation (explicit instruction) type classes. The others had been attending classes utilizing the Total Physical Response method and visual aids (implicit instruction). It should be noted that both classes followed a syllabus which specified the structures and vocabulary which were to be covered. Winitz found that learners in the implicit instruction classes outperformed learners receiving explicit instruction on the grammaticality judgement test.

Even though Winitz is clear in stating at the beginning of his study that "the terms *explicit* and *implicit* refer primarily to language instructional approaches" (p.32), in drawing his conclusions he conflates explicit and implicit *instruction* with explicit and implicit *learning*, assuming that learners receiving explicit instruction learned explicitly and learners receiving implicit instruction learned implicitly. It is then concluded that implicit learning processes are superior to explicit learning processes in developing learners' grammatical competence. Exactly what learners in each instruction group were attending to and becoming aware of during their language lessons is left unanswered. One possibility is that the implicit instruction was more interesting and thus better able to keep learners focused on the task of learning Spanish, and as a result that learners receiving implicit instruction were noticing more than those receiving explicit instruction. Perhaps the use of, for instance, uptake charts following each class session, on which learners write down what they think they have learned, could have helped illuminate what and how much learners were actually becoming aware of.

Long et al. (1998), the most recent study listed in Table 2.0, is actually two studies, one involving Japanese as a foreign language[3] and the other Spanish as a foreign language, of the effects of models and recasts on learners' acquisition of a particular structure. Since recasts are believed to be a useful way to focus on form (Doughty & Williams, 1998), conducting this kind of empirical research on their efficacy is very important. The researchers found some evidence that adults can learn from recasts and that recasts are more effective than preemptive models. Unfortunately, though, the evidence is not quite overwhelming.

In the Long et al. studies (1998), as in the other studies critiqued above, there is no information of how the learners respond to recasts or of what they become aware of when they receive recasts. For example, it seems likely that some learners would recognize recasts as corrections and try to learn from them, while other learners would treat recasts as just some peculiar kind of backchannel. Perhaps debriefing interviews could have been used to assess how learners perceived recasts and what they became aware of as a result of the recasts. This could have shed some light on how recasts may work and any relationship between recasts and awareness. Additionally, had information on how learners did or did not react to recasts been included, it may also have illuminated how learners were perceiving recasts.

It should be noted again that the above studies were not designed to investigate learner awareness or other learner-internal processes. Each of these studies provides useful information on the effects of different kinds of instruction and it is unfair to criticize them for not investigating awareness. In the next subsection, studies which do address learner-internal processes will be critiqued for how well they accomplish this.

Measuring learner-internal processes. Eight of the studies listed in Table 2.1 are on vocabulary acquisition. Five of these include some measure of learner-internal processes. For example, in one study on the effects of premodified versus interactionally modified input (R. Ellis, 1995), uptake charts were used to assess what vocabulary learners in the interactionally modified input group were aware of learning. For some reason, these charts were not used for learners in the premodified input group. In another study (Hulstijn, et al., 1996), a word recognition test was used to assess whether learners were aware of vocabulary that had appeared in the story used in instruction. Such measures indicate that awareness is important during incidental vocabulary acquisition. None of the studies, though, attempt to measure vocabulary acquisition as a function of learner awareness.

The majority of studies in Table 2.1 deal with the acquisition of morphosyntax rather than vocabulary. Of these, the earliest to include a measure of learner-internal processes is Fotos (1993), in which noticing was measured by asking learners to underline "special uses of English" (p. 390). Two other studies (Alanen, 1995; Leow, 1998) used think-aloud protocols to measure noticing. In Fotos (1993) and Alanen (1995), it was found that form-focused instruction resulted in more noticing and more acquisition. In Leow (1998), it was found that learner-centered activities and multiple exposure facilitated noticing.

It is important to know how well different kinds of instruction can facilitate noticing, but in order to investigate the causal role of awareness n SLA, it is even more important to investigate acquisition as a function of learner awareness. In the Alanen study (1995), for example, it was found that learners who focused on target features tended to acquire them. In a different study from that cited above, Leow (1997) compared learners displaying different levels of awareness, a direct test of the noticing hypothesis. This study is not included in Table 2.1 since it does not involve comparison of different kinds of instruction. Using think-alouds collected from elementary learners of Spanish as a foreign language completing a crossword puzzle designed to draw their attention to the spelling changes of stem-changing verbs, Leow distinguished three levels of awareness: lack of awareness, awareness unaccompanied by a rule, and awareness with a rule. Leow found these different levels even though all learners were performing the same task, indicating that it cannot be assumed that all learners in a given instruction group learn in the same way. Leow then compared learners at these different levels on post-crossword puzzle recognition and production tasks and found that greater awareness during the task was predictive of more recognition and greater productive accuracy.

Finally, Robinson (1995a, 1996a, 1996b, 1997a) has compared the awareness of learners in implicit, incidental, rule-search, and instructed conditions and has also (Robinson, 1995a, 1997a) attempted to compare the quality of learning displayed by learners with more or less awareness. In order to assess learners' awareness levels, Robinson gave each learner a debriefing questionnaire immediately following completion of computerized instruction. Learners were asked whether they had been looking for rules, whether they had noticed rules, and whether they were able to verbalize rules. Learners who answered the last question positively were also asked to attempt this verbalization and were classified as able-to-verbalize if they could give a rule, rule fragment, or relevant exemplar.

Robinson was able to make some interesting findings by looking at the awareness of learners in different instruction conditions. First it was found that some learners in all four instruction groups, even the implicit and incidental groups, answered that they were looking for rules, that they did notice rules, and that they were able to verbalize rules. A chi-square analysis comparing the frequencies of learners in different conditions who answered these questions positively revealed that only at the level of looking for rules was there a significantly greater number of learners in the rule-search and instructed conditions. Even so, over one third of learners in the implicit condition and over one half of learners in the incidental condition stated that they had been looking for rules. Also interesting is that not all learners in the rule-search and instructed conditions answered that they had noticed, had been looking for, or were able to verbalize rules. In fact, over one fifth of learners in the rule-search condition stated that they had not been looking for rules, despite instructions to do so. As in Leow (1997), this shows quite clearly that it should not be assumed that learners necessarily learn as researchers intend them to. Learners receiving implicit, incidental, or similar instruction may be intentionally trying to learn, while learners receiving more explicit instruction may be failing to make use of it. Similar results were found in Robinson (1997b), which compared learners in implicit, incidental, enhanced (rather than rule-search), and instructed conditions.

Robinson (1995a, 1997a) also compared the learning success of learners in each condition who were aware at the levels of noticing, looking for rules, and able to verbalize rules with those who were not. Learners in four different instruction conditions, described above, had received instruction on or exposure to an easy pedagogical rule and a hard pedagogical rule. By comparing the success of learners who reported awareness with those who did not, Robinson found that learners in the implicit condition who reported that they had been looking for rules were significantly more successful in learning both the easy and the hard rule than learners in the same condition who had not been looking for rules. The same was true for implicit learners who were able to verbalize rules. In addition, learners in the rule-search condition who were able to verbalize rules were more successful on the hard rule than those who were not. This indicates not only that learners in supposedly implicit, unconscious learning conditions may actually be attempting to learn intentionally, but also that those learners who do attempt this may be more successful as a result of their higher level of awareness.

Importance of Learner-Internal Processes

The studies described above indicate that a) explicit instruction, learner involvement, and multiple exposure facilitate noticing, b) learning occurs with awareness, which, of course, results from noticing, and c) learner behavior is not determined solely by the type of instruction they receive. This last fact highlights the importance of investigating learner-internal processes since, for example, learners who receive some form of non-explicit instruction may nevertheless be trying to learn explicitly while other learners receiving explicit instruction may not. Learner behavior will also influence what is noticed. Evidence for this dissociation between type of instruction and learner behavior was found in the studies by Leow (1997) and Robinson (1995a, 1996a, 1996b, 1997a, 1997b) discussed above.

Further evidence for this dissociation comes from DeKeyser (1995). In an effort to investigate *implicit-inductive* and *explicit-deductive* language learning in a tightly-controlled atmosphere, DeKeyser conducted an experiment in which two groups of learners received computerized instruction in the use of a miniature linguistic system. By using such a system, DeKeyser was able to control for prior learning (there could be none) while simultaneously imitating natural language learning by using a system which could, in a limited way, be used to communicate information. While DeKeyser found a benefit for explicit instruction for categorical rules, as opposed to prototypical patterns, from the point of view of the present paper one of the most interesting elements of the DeKeyser study was his use of retrospection. Learners were asked, among other things, whether they had been looking for grammar errors and how much they had been thinking about grammar. Learners also reported that they were able to understand most of the language used in the experiment. In addition, learners in the implicit-inductive group "were asked whether they had suspected there was a system of grammar rules underlying the sentences they had seen" (p. 396). Seventeen out of the twenty-one learners in this group said they had. This is interesting because it illustrates the importance of not assuming that learners will behave as researchers intend. It is possible that learners receiving implicit instruction will end up explicitly focusing on the targets. As DeKeyser points out, learning strategies can override distinctions drawn by researchers.

While DeKeyser (1995) found that learners receiving implicit instruction may actually be trying to learn explicitly, DeKeyser (1997) is also sensitive to the fact that learners cannot be assumed to be aware merely as a result of having received explicit instruction. In this study, DeKeyser compared the comprehension and production abilities of groups of learners who had extensively practiced either production or comprehension after receiving explicit instruction on an artificial language. It was found that production practice led to improvement only in production and, in contrast to VanPatten and Cadierno (1993) and Cadierno (1995), that comprehension practice led to improvement only in comprehension. Importantly, before beginning practice, learners had to reach a criterion level of awareness on a metalinguistic test, thus controlling for the fact that explicit instruction does not guarantee awareness.

Because type of instruction is dissociated from type of learning and does not determine how learners will approach the task of learning a language, it is important for researchers attempting to investigate different types of learning, such as explicit, implicit, incidental, or intentional learning, not to assume that type of instruction determines what learners do. Also, in investigations of the role of awareness in SLA, it is important not to equate explicit instruction with learner awareness or non-explicit instruction or exposure-only with lack of awareness.

Measuring Learner Awareness

Several of the effects of instruction studies discussed above are admirable in that they include some measure of learner awareness. For studies of vocabulary acquisition, such measurement seems relatively straightforward. Researchers can easily make use of uptake charts or word recognition tests. Things are more complicated, though, when researchers attempt to measure awareness of some kind of rule, as needs to be done in studies of the acquisition of morphosyntax. Research on awareness and learning in the field of cognitive psychology can be very instructive for researchers in the field of SLA attempting to investigate awareness. Three examples of such research will be discussed here.

One example is Curran and Keele (1993), discussed in the previous chapter, in which sequential learning under single-task and dual-task conditions by learners who had and had not been informed of the sequence was compared. Rather than assuming that learners who had not been informed of the sequence were unaware, the researchers used interviews to probe these learners' awareness and then divided them into more-aware and less-aware groups, recognizing the impossibility of categorically stating that less-aware learners were completely unaware. While the researchers' claims for non-attentional learning are not very convincing, their use of a measure of awareness to discriminate among more-aware and less-aware learners, who had received the same instruction, is commendable and should be emulated in SLA.

A very different approach was taken in Clark and Squire (1998). In this study, the researchers compared delay conditioning and trace conditioning by normal participants and participants with hippocampal lesions. In delay conditioning, a conditioned stimulus is presented slightly before the onset of an unconditioned stimulus and the two stimuli are terminated simultaneously. In trace conditioning, a conditioned stimulus is also presented slightly before the unconditioned stimulus, but the conditioned stimulus is terminated slightly before the onset of the unconditioned stimulus. Amnesic patients with hippocampal lesions are claimed to be incapable of developing conscious awareness, so any kind of learning which they are capable of must not rely on such awareness. In addition, awareness measures were used to discriminate between more-aware and less-aware normal participants. All participants displayed delay conditioning, but amnesic participants and less-aware normal participants did not display trace conditioning while more-aware normal participants did. They concluded that trace conditioning was dependent on the hippocampus and awareness while delay conditioning was dependent on neither.

The methodology used in this study would probably be impractical for SLA researchers to replicate in its entirety, due to a shortage of people with hippocampal lesions attempting to learn a second language, but the type of sensitive measure used to discriminate more-aware and less-aware normal participants probably can be applied in SLA research. Immediately following the experiment, normal participants answered a list of 17 true-false questions on the relationship between the conditioned and unconditioned stimuli. They were not allowed to change their answers. Participants who answered correctly at above chance (at least 13 correct answers) were classified as more-aware. This is an approach that could be used in SLA research to create fairly sensitive measures of learner awareness for use in postinstruction debriefing sessions.

The difficulty of investigating learner awareness is apparent in the research on rule-based and instance-based learning conducted by Anderson, et al. (1997). While in the first experiment the researchers asked participants whether they were using rules or examples to base their decisions on, the researchers were dissatisfied with participants' ability to give this information and stopped asking participants in the other experiments reported in this article.

Developing adequate measures of awareness is critical for research on the relationship between awareness and learning, whether in SLA or in cognitive psychology. Whether measures of awareness are adequate can be determined by how well they meet the information criterion and sensitivity criterion (Shanks & St. John, 1994) described in the previous chapter. These criteria were developed for studies of learning in the field of cognitive psychology, but they are applicable to and should be applied in SLA research on the relationship between language learning and awareness.

Conclusion

It is argued by many SLA researchers and believed by a lot of people in general that SLA occurs without awareness of what is being learned. However, there exists no research on the relationship of awareness to SLA which meets the criteria described in Chapter One. For example, Robinson (1995a, 1996a, 1996b, 1997a) admits that the awareness measure used in his research—asking learners whether they had noticed rules, had been looking for rules, and were able to verbalize rules—is not very powerful and does not meet Shanks and St. John's criteria. In fact, though, Robinson's work is a step forward in SLA, as many studies use no awareness measure at all, assuming that how learners behave will be determined by their membership in a particular instruction group. The use of concurrent think-alouds by Leow (1997) and Alanen (1995) to measure learner awareness can also be seen as a step forward, especially since it is an attempt to investigate awareness at the point of learning rather than soon after, but it probably does not meet either of the criteria listed above since learners are probably unable to simultaneously explicate *all* that they are aware of and complete a pedagogical task. Meeting Shanks and St. John's criteria may be an extremely difficult task, but, if the relationship between SLA and awareness is to be empirically investigated, measures of awareness must be used which at least attempt to meet them. The measure of awareness used in this study will be described in the next chapter.

Chapter Three
The Present Study--Experiment One

Introduction

The purpose of the present study is the investigation of the relationship between awareness and learning. Put simply, the researcher expects that more awareness will accompany, possibly even cause, more learning. Of particular interest is the relationship between awareness and learning among learners receiving incidental instruction, who have been instructed to process meaning. Also of interest is whether a comparable number of learners receiving incidental instruction display awareness when compared to learners receiving explicit instruction.

The different strategies of learners receiving identical instruction is also of interest. Various studies of classroom language learning (e.g., Allwright, 1984; Slimani, 1989, 1992) have shown that learners in language classrooms do not necessarily focus on what teachers intend them to. The experimental studies cited above, especially Robinson (1995a, 1996a, 1996b, 1997a), Leow (1997), and DeKeyser (1995), have also demonstrated that learners have some control over how they approach a learning task. Again, it cannot be assumed, for example, that learners receiving incidental instruction will not focus their attention on and become aware of form.

According to Krashen (1994), second language acquisition, as opposed to language learning, occurs both incidentally and unconsciously. It is claimed to be incidental in the sense that learners are concerned with meaning, not with form. It is claimed to be unconscious in the sense that learners are not aware of the formal features that they are acquiring. However, as can be seen in the Robinson (1995a, 1996a, 1996b, 1997a), Leow (1997), and DeKeyser (1995) studies cited above, while acquisition may be incidental in the sense that learners can learn while carrying out tasks which focus on meaning, such incidental acquisition is not necessarily unconscious. Learners may very well become aware of formal features while working on a meaning-oriented task. Instead of being unconscious, successful incidental learning may involve, at the very least, learners consciously noticing formal features of the language in the process of using and comprehending the language.

This study aims to investigate to what extent learners receiving two different kinds of instruction become aware of a target feature during instruction. It then draws comparisons between the performance of those learners who display greater awareness of the target with those who do not. If some awareness is necessary for learning, and if different levels of awareness can be captured by the debriefing session which is used in this study, then those learners who display greater awareness should outperform those who do not.

Finally, the roles of instance-based and rule-based learning are of somewhat minor interest in this study. Does instance-based learning play a more important role than rule-based learning in the learning of subjects receiving incidental instruction? Does instance-based learning play a more important role in the learning of less-aware learners? Does instance-based learning still play a major role in the learning of learners who have received explicit instruction? Does it still play a major role in the learning of more-aware learners?

Experiment One

As in many effects of instruction studies, this study compares the learning, as measured by a pretest and a posttest, of groups of learners receiving different types of computerized instruction on some feature of English. There are two *instruction* groups: an incidental instruction group and an explicit instruction group. These are called *instruction* groups rather than *learning* groups since some learners in the incidental instruction group may nevertheless try to learn intentionally, while some learners in the explicit instruction group may not make use of the explicit instruction they receive. Both groups are given a computerized grammaticality pretest immediately prior to instruction and a similar posttest immediately following instruction. For the incidental group, instruction consists of being asked to respond true or false to statements which appear on the computer screen. Learners in this group are given feedback on the correctness of their responses. This instruction is incidental in that learners are asked to process meaning, not to focus on the target feature. For the explicit instruction group, instruction consists of rules of thumb plus examples, which learners are instructed to memorize. This is explicit in that learners are asked to focus on the target feature. Learners in all groups control the pace of both the grammaticality tests and the instruction.

Half the items (40) on the posttest are identical to the items used during instruction and will have also appeared on the pretest. The other half are novel items. All the items used during instruction are on the posttest and on the pretest.

Debriefing sessions conducted in the learners' first language, Japanese, will be conducted immediately following completion of the posttest. The data collected in these sessions will be used to divide learners into a more-aware *learner* group and a less-aware *learner* group. (See below for details.) The learning success of these two groups will then be compared.

It is expected, based on the empirical evidence cited above (Doughty, 1990; DeKeyser, 1995; Robinson 1995a, 1996a, 1996b, 1997a, 1997b), that the explicit instruction group will outperform the incidental instruction group. It is also expected, based on the empirical work of Leow (1997) and Robinson (1995a, 1997a), as well as on the noticing hypothesis (Schmidt, 1990, 1993, 1994b, 1995), that the more-aware learner group will outperform the less-aware learner group. Finally, while it seems that learners in the explicit instruction group will be more likely to fall into the more-aware learner group, it is expected that there will be no simple one-to-one correspondence between instruction groups and learner groups. In other words, there will be learners from the explicit instruction group in both the more-aware and less-aware learner groups, as well as learners from the incidental instruction group in each of the learner groups.

Because of the possibility that the instruction may not result in any long-term learning, a small-scale delayed posttest will be given to a subset of learners in the more-aware learning group and the less-aware learning group. Also, since grammaticality judgement tests may only tap learners' monitoring ability (Krashen, 1981, 1982, 1994; Paradis, 1994), this test will involve using English and the target feature in actual communication.

This study addresses several research questions. The first three research questions concern the two types of instruction, incidental and explicit:

1. To what extent will learners receiving computerized instruction, either incidental or explicit, on the use of definite and zero articles with place names

show improvement on this use of articles, as measured by correct scores on a computerized grammaticality judgement test?

2. Will the explicit instruction group display greater improvement than the incidental instruction group, as would be expected based on the empirical evidence cited above?

3. As measured by correct scores, to what extent will learners in the explicit instruction group be better able to generalize from examples used on the pretest and in instruction to novel items on the posttest, as studied in Robinson (1997b)?

The next four research questions concern the more-aware and less-aware learner groups:

4. Will more learners in the explicit instruction group than in the incidental instruction group fall into the more-aware learner group?

5. Will learners in the more-aware learner group display greater improvement than learners in the less-aware learner group, as predicted by the noticing hypothesis and as found in Robinson (1995a, 1997a) and Leow (1997)?

6. Will learners in the more-aware learner group be better able to generalize from examples on the pretest and used in instruction to novel items on the posttest, as measured by correct scores?

7. Will learners in the more-aware learner group continue to display greater learning on a delayed, non-computerized, communicative posttest?

A final research question concerns learner response times on previously presented items and novel items on the posttest:

8. Will there be differences in response times to previously encountered items and novel items?

Developing an adequately sensitive measure of awareness is imperative for this study. The measure of awareness used in Robinson (1995a, 1996a, 1996b, 1997a) involved asking learners whether they were looking for rules, whether they had noticed any rules, and whether they could verbalize any rules. Verbalizations were accepted if they contained a relevant exemplar or rule. As Robinson admits, this method of measuring awareness was very crude. For the present study, a much more sensitive measure is desired. Leow used comments about spelling changes to classify learners as aware and any comments about spelling change rules to classify learners as aware with a rule. As described above, the comments were all collected by asking learners to think aloud as they completed the task. This seems to be an adequately sensitive measure, in that it probes awareness at the time of learning, but it was decided in this study that think-alouds could interfere with the learning task by, for example, directing learners' attention to the target feature (Shanks & St. John, 1994) or by using up learners' attentional capacity and preventing them from being able to attend to learning (Jourdenais, 1996). It was also decided that the number of subjects anticipated for this study (60) would render concurrent think-alouds impractical. Some of the other methods of assessing noticing and awareness discussed in the previous chapter, such as underlining special uses of English (Fotos, 1993) or a vocabulary recognition test (Hulstijn, et al., 1996) are not appropriate for this study since it involves computerized instruction and is not concerned with vocabulary acquisition.

In a pilot study, conducted by the researcher, comparing the learning of more-aware and less-aware learners who had received fairly explicit classroom instruction on the use of the English article system, learners were asked to write down, in English, what they had learned about the use of articles immediately after the posttest. They were told that they could write rules, examples, or both. Two serious methodological problems prevented the researcher from being able to compare learners at different levels of awareness. First, learners' willingness to write in detail in their second language about the knowledge they had gained varied greatly from learner to learner. Some learners wrote a great deal, including both rules and examples. Others just made a few, sometimes illegible, notes. This may have reflected lack of awareness, but it may also have been caused by some learners' general lack of verbosity or by the fact that they were asked to write in English rather than in their first language. Obviously, this approach did not meet the sensitivity criterion of Shanks and St. John's (1994). Second, there was too great a variety of learning targets. Evidence of learners' awareness of certain specific uses of articles in English was obscured by the variety of uses for articles which were covered in the instruction.

It was determined that if debriefing sessions are going to be effective in measuring learners' awareness, they must be conducted orally so that the researcher is able to probe the learners' awareness and force them to divulge as much information as possible. It was also determined the debriefing interviews should be conducted in the learners' first language, in this case Japanese, a language in which the researcher is proficient. Finally, it was decided that target features should be more focused.

As pointed out by Shanks and St. John (1994) and Robinson (1995a, 1996a, 1996b, 1997a), asking learners to give verbal reports tends not to be a very sensitive measure of awareness. However, it was decided that such reports would be the only practical measure that could be used in this study. Therefore, it was decided that the verbal reports must be made as sensitive as possible. First of all, the reports were collected immediately following the posttest in order to avoid the problem of learners being unable to recall what they were aware of during learning (see Green & Hecht, 1992, for an example of this problem). Next, as described above, the debriefing uses probing questions (e.g., *What else did you learn? Are you sure you didn't learn anything* *else?*), followed by more direct questions (e.g., *Do you use* the *with mountains? With mountain chains?*), to explore the extent of learners' awareness more exhaustively than would be possible by simply asking learners what they were aware of (Shanks & St. John, 1994). The debriefing continued until the researcher was satisfied that enough information had been collected about how aware learners are of the twenty uses of the target feature (see below) that were used in this study. Similar to Clark and Squire (1998), learners are classified as more-aware if they perform above chance, knowing correctly at least 15 out of 20 uses. Finally, again as described above, the debriefing was conducted in Japanese, the first language of all the learners who participated in this study. While still rather crude, it was anticipated that these debriefings would be sensitive enough measures of awareness to discriminate between more-aware and less-aware learners. This level of sensitivity is deemed adequate since the study does not rely on any necessity to identify learners with absolutely no awareness.

One objection that could be raised about this study is that it is rather obvious that learners who display more awareness during the debriefing will have performed better on the posttest. After all, the posttest and the measure of awareness seem to be measuring the same thing—whether learners know how to use definite and zero articles with place names. However, in the debriefing sessions, the researcher only asks questions to push learners to reveal their awareness of rules—in other words, their explicit knowledge. The researcher does not ask learners to apply their knowledge, whether it be explicit or implicit, of the use of articles with place names. If learning can truly occur without awareness, then it should be possible to gain knowledge of *how* definite and zero articles are used without actually being aware of the knowledge. In this study, though, it was not expected that learning without awareness would occur.

A second objection to the methodology of this study is that it is not all that difficult for learners in the incidental instruction group to figure out what the target features are. For example, Reber (1989) states that the stimulus domain must be sufficiently complex in order for implicit learning to occur. In this study, though, the pretest should alert many learners in this group that the instruction concerns the use of definite and zero articles with place names. However, one of the main purposes of this study is to observe how the learners themselves choose to focus their attention. Some learners in the incidental instruction group may follow instructions and focus on meaning, but others may also focus on the target feature. In order to compare the learning of incidental instruction group learners who choose to focus on forms with the learning of those who do not, the more forms-oriented learners must focus on the right thing. If learners followed a strategy of attending to forms but direct their attention to something other than the target feature, it would be impossible to discriminate them from less forms-oriented learners. For this reason, it must be possible for these learners to realize what the instruction is all about.

Related to this, some learners in the explicit instruction group may follow instructions and focus on the target feature, but others may just try to get through the experiment as quickly as possible. It must be possible for these learners, whose personal preference may be to ignore formal features of the language, either to fail to learn or to forget what they have learned. For this reason, a distracter task followed the instruction.

The Target Feature

It was decided that the target feature of instruction must meet four criteria, in addition to being a feature of the English language, with large scope and high reliability (Hulstijn & De Graaff, 1994), which learners of English as a Second Language (ESL) would benefit from acquiring. The first criterion was that it be outside the scope of Universal Grammar (UG) (Hulstijn & De Graaff, 1994). The second was that the use of the feature be explainable by a set of simple rules of thumb (Hulstijn, 1995). The third was that the feature would be complicated enough to be a challenge for learners to master (Hulstijn & De Graaff, 1994). The fourth was that the feature would allow prior learning to be controlled for. The linguistic feature which was eventually chosen for this study was the use of either the definite article or zero article with place names (e.g., *the Nile River, Mount Everest*).

It is difficult to determine for sure whether or not a given feature is within or outside the scope of UG. Of course, the issue of whether UG is even relevant to SLA is certainly not settled (Anderson, 1995), and there are those who argue that even first language acquisition can be explained without resorting to U.G. (e.g., Anderson, 1983; Winter & Reber, 1994). However, it is important, for the purposes of this study, that the target feature be outside the scope of UG since, if UG plays a role in SLA, the innate nature of UG may obscure the role of awareness in learning the feature. Also, this paper is not concerned with any possible relationship between UG and SLA, between UG and L2 instruction, or between UG and awareness. Therefore, it was thought important to avoid UG altogether. While there seems to be a lack of research or theory indicating that the use of definite and zero articles with place names is outside UG, there is a similar lack of evidence that it falls within the scope of UG. This particular use of articles is rather arbitrary and it would seem safe to assume that UG does not play a role.

In this study, in order to compare more-aware and less-aware learners as well as learners in different instruction groups, it is hoped that some learners in each instruction group become aware of simple rules for this use of articles. It is important that learners in the explicit instruction group, who are presented with rules and examples during instruction, be able to understand the rules so that they can apply them when they are tested. It is also important that those learners in the incidental instruction group who focus their attention on article use during instruction have the chance to figure out what the rules are. Therefore, the criterion that the target feature be explainable by a set of simple rules of thumb was adopted. The use of definite and zero articles with place names meets this second criterion. Even though much more complicated linguistic explanations of this target feature exist (Hewson, 1972; Kaluza, 1981; see below), when to use the definite article and when to use the zero article can be described by extremely simple rules: *Use "the" with rivers. Do not use "the" with single mountains.* These rules not only have simple syntax, but, they can also be expressed without using any metalinguistic terminology. To the extent that learners know vocabulary such as *river* and *mountain*, relatively low proficiency in English should not prevent them from being able to understand these rules. They also seem simple enough for some learners, given enough input, to figure out on their own.

The third criterion, that the feature be complicated enough to be a challenge for learners to master, may seem to contradict the second. However, the second criterion concerns the simplicity of the rule statements, while the third criterion concerns the complexity of how the feature is used. If the feature's use is too simple, it would be too easy for learners to master and differences in the quality of learning would be difficult to detect. For example, it would seem that the use of the two forms of the English indefinite article (*a/an*) is very simple, based only on the initial sound of the following word. Explained as a pedagogical rule, one rule would be sufficient: *Use* an *before a vowel sound and* a *elsewhere.* It would also probably be simple for any learner who noticed this feature to master it rather quickly, at least in speaking if not in writing.

The system for using definite and zero articles with place names, though, is quite complex and potentially quite confusing. Master (1990) argues that this use of articles appears to be arbitrary and cannot be explained by pedagogical rules developed to teach other uses of articles. For example, *the* is used with mountain chains and island groups (*the Rocky Mountains, the Hawaiian Islands*), but not with the names of individual mountains and islands (*Mount McKinley, Oahu*). (*The Big Island*, used in the state of Hawai‘i to refer to the island of Hawai‘i, is an exception.) *The* is used with the names of universities that contain the word *of* (*the University of Hawai‘i*), but not with the names of universities that do not (*West Virginia University*). (*The Ohio State University* is a notable exception.) Also, most country names take zero article (*Mexico*), but some take *the* (*the United States of Mexico*). Attempts to find systematicity in this use of articles (see below) are not very successful at removing this complexity. Master (1994), in a study on the effects of systematic instruction on the use of articles over a ten-week quarter, seems to have recognized this lack of systematicity and omitted the teaching of articles with place names.

The existence of prior learning is a common problem in research involving a natural language, especially such a widely studied language as English. It would probably be impossible to eliminate completely the possibility of prior learning in investigations of ESL learners. However, it should be possible to find a feature of English which learners do not seem to know much about, either explicitly or implicitly. For example, Robinson (1995a, 1996a, 1996b, 1997a) controlled for prior learning by choosing learners on the basis of incorrect grammaticality judgements of grammatical sentences based on the easy and hard rules used in his study. Similarly to Robinson, grammaticality judgements were used in this study to control for prior learning—learners who performed above chance on the pretest (52 or higher out of 80, .01 probability) were removed from the study. Also, it has been the researcher's experience that even learners who have been studying English for over six years in formal contexts, in addition to living in English-speaking countries, rarely know much at all about the use of articles with place names and that many pedagogical grammars do not deal with this feature.

The use of definite and zero articles with place names adequately meets the four criteria listed above. In addition, it seems to be of large scope and fairly high reliability (Hulstijn & De Graaff, 1994; Hulstijn, 1995), even though there are exceptions to how articles are used (*the Gambia* (a country as well as a river in Western Africa), *the Ohio State University*). Also, when the researcher has taught this feature to ESL students, some of them have commented on the usefulness of the instruction. (No one has ever commented on its uselessness.) Therefore, it was decided that ESL learners would benefit from acquiring this feature.

Even though the researcher has often been told by learners that instruction on this use of articles is useful, it should be noted that there also seems to be negligible communicative content in the use of zero and definite articles with place names. It seems unlikely that omitting the definite article would cause confusion, and, indeed, in certain situations, such as on maps or in a hotel's promotional literature, the definite article is often omitted. This is another sense in which this use of articles is different from other uses of the English article system. Even though Master (1994) argues that correct article usage only becomes important in writing and that in speaking incorrect usage is unlikely to lead to communicative problems, evidence from studies in language processing reviewed in Anderson (1995) indicates that articles, particularly the definite article, convey more meaning than they are often given credit for, even in speaking. This does not seem to apply, though, to the arbitrary use of articles that is the focus of this study.

There seems to be very little research in the field of linguistics into this use of zero and definite articles. However, one linguist, Hewson (1972), has attempted to develop semantically-based criteria for choosing the definite or zero article, and this attempt in itself is indicative of the complexity of the system. According to Hewson, the majority of place names take zero article. However, plural or collective place names take the definite article, possibly in order to avoid suggesting that a plural place name is a common noun unlimited in number. For example, calling *the Hawaiian Islands* just *Hawaiian Islands* would suggest that there exists a virtually unlimited number of things called *hawaiian islands,* just as the phrase *the cups* refers to a finite set of cups while *cups* is potentially infinite. This use of the definite article includes not only such obviously plural entities as mountain chains and island groups, but also things such as unions and federations. *The United States of America* takes the definite article because it is seen as a federation of several states. This explanation seems to make some sense, but it is not clear why, for example, *the Republic of Ireland* should take the definite article. Also according to Hewson (1972), the names of places which are hard to define or do not have easily definable boundaries also take the definite article. For example, it is hard to define what exactly is a desert and what is just a fairly dry place, so *the Sahara Desert* takes the definite article. However, as Celce-Murcia and Larson-Freeman (1983) point out in a discussion of Hewson's system, it is not clear why, for example, oceans take the definite article (*the Pacific Ocean*), while bays take the zero article (*Kaneohe Bay*). Also, there are many exceptions to Hewson's system (e.g., *The Hague)*.

Kaluza (1981) has also attempted to find sytematicity in the use of definite or zero articles with place names. In this system, plural names—island chains, groups of lakes, mountain chains--are seen as taking *the*. Compound forms, made up of more than one word, take the definite or zero article depending on the head noun. When the head is arbitrary, meaning that it has no semantic content other than being a name (*Britain*), or once had semantic content but has become disassociated with that content over time (*Newfoundland*), then the compound noun takes zero article, just as simple arbitrary names take zero article. For example, the head of *Great Britain* is *Britain*, which is arbitrary, so *Great Britain* takes zero article. When, though, the head is a countable noun, such as *kingdom*, the compound takes the definite article, making it explicit exactly what is being referred to. An example is "*the United Kingdom* (which Kingdom (sic)? The kingdom that is united under one monarch, now Elisabeth II)" (p. 65). This distinction would seem to account for names such as *the Republic of Ireland* (which republic? The republic founded in the place called Ireland.) It would also account for differences such as *the University of Hawai'i* and *West Virginia University*. In the former, the head noun is *university*, a countable noun, so the definite article is necessary. In the latter, the head noun is *West Virginia*, an arbitrary name, so the compound name takes zero article.

These descriptions of the use of zero and definite articles with place names are interesting from a linguistic point of view, but it seems doubtful that they correspond, in a psychologically real way, with the knowledge of native speakers and native-like ESL speakers. Such knowledge seems more likely to be based on memory for specific instances, which may be either retrieved directly or used by analogy. In this study, explicit instruction and measurement of learners' awareness will not be based on the descriptions given above, but, rather, on a list of rules of thumb for article use, which can be derived from specific examples.

Both Hewson (1972) and Kaluza (1981) provide lists of place names that take and do not take definite articles. The researcher used these lists, lists provided in a widely used pedagogical grammar, *Grammar Dimensions 2* (Riggenbach & Samuda, 1997), and his own intuition to choose the place names which would be used in the study. Ten types of place names which take zero article were chosen: schools (*Kaimuki Elementary School*), most countries (*Canada*), universities without the word *of*, (*West Virginia University*), parks (*Kapiolani Park*), single mountains (*Mount Fuji*), parts of continents (*Central Africa*), hospitals (*General Hospital*, *the Queen's Medical Center* is a notable exception), single lakes (*Lake Geneva*), bays (*Hanauma Bay*), and single islands (*Oahu*). Also chosen were ten types of place names which take the definite article: tourist attractions (*the Washington Monument*), regions not including a continent or country name (*the East Coast*), island chains (*the Hawaiian Islands*), hotels (*the Hilton Hotel*), mountain chains *(the Alps*), certain country names (*the United States*), universities with *of* in the name (*the University of Hawai'i*), rivers (*the Yellow River*), peninsulas (*the Korean Peninsula*), and associations (*the Young Women's Christian Association*). (Associations, admittedly, are a rather marginal member of the place name category.) A paper and pencil pilot grammaticality test (see below) also included city names and continent names. However, these were removed because learners scored almost 100% on judgements of these items. (See Appendix A for a complete list of items that were used on the pretest and posttest and during instruction. See Appendix B for a list of the 20 rules of thumb.)

Method

Participants. Sixty young-adult Japanese ESL learners studying at the Hawai'i English Language Program (HELP), Intercultural Communications College (ICC), and other English language schools in Hawai'i participated in this study[4]. They were paid five dollars each for their participation. Learners were randomly assigned to one of the instruction groups. No intact classes were used. All learners had studied English for at least six years in formal foreign language classrooms. In order to control for prior learning, learners who scored above chance (52 or higher) on the pretest were removed from the study. All learners are intermediate in that they are able to understand the task directions provided on the computer screen, but not advanced enough to command already this use of articles.

Materials. All materials, except the delayed posttest, were computerized using PsyScope software (Cohen, et. al, 1993).

A grammaticality judgement test containing 80 items was used to pretest learners. Each item consisted of the question, “Which is correct?” followed by two choices numbered one and two. For example:

Which is correct?

1. Nile River
2. the Nile River

The items appeared on the computer screen one at a time. Learners responded by pressing one or two on the computer keyboard. The next item appeared on the computer screen immediately after the previous item had been answered. No feedback on correctness was given. PsyScope recorded learners’ responses and response times. In order to allow learners to practice responding to items in this format, the pretest was preceded by ten addition problems such as:

Which is correct?

1. 4+5=9
2. 4+5=20

In order to develop a test consisting of items which would be unknown to the learners, a 99 item pencil-and-paper pilot test was administered to ten predominantly Japanese (seven Japanese, two Korean, and one Chinese) learners who were believed to be of approximately the same proficiency as the learners who would participate in the study. Item facility (IF) was then calculated for each item, and items with a high IF were examined in an effort to find categories of items that learners seemed to know already.

The IF figures were between 0.40 and 0.60 for the majority of items, indicating that learners were guessing on most.

After examining items with a high IF, the researcher found that city names (no article), store and restaurant names (no article), continent names (no article), and certain country names presented little or no difficulty for the learners. Learners generally had no problem with country names such as France, Mexico, and Canada. However, they had much more difficulty, operating at around chance level, with country names which take the definite article, such as *the United States*, *the Republic of Ireland*, and *the United Kingdom*. Since this contrast in the use of zero article and the definite article with the names of countries seemed to be difficult for learners, it was decided not to remove country names from the test. Also, while continent names seemed to be very easy, names such as *Central America* and *South Asia* were much more problematic. Therefore, while continent names were removed from the test, this latter category of name was retained. Finally, city names and store names were removed from the test. All the item categories which were removed were ones that take the zero article.

Developing a reliable test was a top priority. However, since the learners who took the pretest operated at around chance level, and since this meant that learners tended to answer about half the items correctly and that each item tended to be answered correctly by about half the learners, the K-R20 reliability estimates for the pilot test were extremely low. Since the major part of the K-R20 formula is one minus the sum of item variance (IV) divided by the total test variance, if the sum of IV is almost the same as the total variance, the K-R20 reliability estimate will be very close to zero. If the sum of IV is a little higher than the total variance, the K-R20 estimate will actually be negative. On the pilot test, IFs tended to be around 0.5, yielding IVs of 0.25, the highest possible IV, while total test variance was rather low since all learners performed around chance level. The result was a negative K-R20 estimate of –0.063. In order to see whether the test would yield a better reliability estimate when all learners were not performing at chance level, a revised 80-item test was administered to the same ten learners who had taken the initial pilot test, plus one additional Japanese learner, after they had received instruction on the use of articles with place names. This time, learners performed at much higher than chance levels, and the K-R20 reliability estimate was 0.733, a fairly high level of reliability. In addition, the even-odd correlation was 0.711 and the full test reliability, using the Spearman-Brown Prophecy formula was 0.831. It was expected that reliability estimates on the computerized version of the pretest would also be quite low since only learners who perform around the level of chance were to be included in the study.

Incidental instruction consisted of 40 statements, each containing one of the items used on the pretest. For example: *The Nile River is the longest river in Africa.* Each statement appeared on the computer screen one at a time. Half the statements were true and half were false. Learners were instructed to decide whether each statement was true or false and then to press the space bar for the correct answer. They then pressed the space bar again to continue. This instruction was called incidental since learners were instructed and required by the task to respond to content, not to focus on formal features. An example of the target feature, though, appeared in each statement.

Explicit instruction consisted of 20 rules of thumb presented with an example. Each rule of thumb appeared on the screen with exactly one example. Learners were able to view the rule and example as long as they pleased and pressed the space bar to continue. They were instructed to memorize each rule and example. After all 20 rules had been presented, the learners viewed, one at a time, 20 more examples, one for each rule. Again, they were instructed to memorize each example and to press the space bar to go on to the next example. There were a total of 40 items presented as examples of the rules, two items for each rule.

The posttest followed the same format as the pretest. It was preceded by one practice item similar to the items used before the pretest, except that this practice item involved subtraction rather than addition. The posttest contained the 40 items from the pretest that had been used in instruction and 40 novel items. The 40 novel items were all based on the rules of thumb used during the explicit instruction. The K-R20 reliability estimate is reported in the results section below.

Learners performed a distracter task following the instruction. On the distracter task, learners were presented with ten pairs of sentences, one pair at a time. They were instructed to press *s* if they thought the two sentences in a pair had the same meaning and *d* if they thought the two sentences had different meanings. They received feedback on their choice. The feedback consisted of the word *right* printed in blue letters or the word *wrong* printed in red letters. The computer then instructed them to press the space bar to continue.

A subset of learners, five in the explicit instruction group (three more-aware and two less-aware) and five in the incidental instruction group (all less-aware), took part in a non-computerized delayed posttest. On this test, learners were asked questions such as, *What is the longest river in the world?* All questions were asked and answered orally, in a relaxed atmosphere with snacks and drinks, in order to simulate as closely as possible a natural conversation. (See Appendix C for a list of questions used on the delayed posttest.) All conversations were recorded and transcribed. Based on the transcriptions, frequencies of correct and incorrect use of articles with place names were counted for more-aware and less-aware learners. This was not too difficult since each place name requires either a definite or a zero article. For example, if they had answered the above question with *the Nile*, their answer would have been coded as correct. If they had answered it with just *Nile*, their answer would have been coded as incorrect. If an answer had been factually inaccurate (*the Kanawha*) but their article use had been correct, it would have been coded as correct. If an answer had made absolutely no sense—they had answered the above question with *the East Coast*—then it would have been removed from the analysis. Only a subset of learners was used for the delayed posttest since it was impossible to schedule a time to take the delayed posttest with all learners and because it would have been impractical to transcribe and analyze this type of test for 60 learners.

Assessing learner awareness. As described above, learner awareness was assessed through oral debriefing sessions, in Japanese, conducted immediately following the posttest. Also as described above, learners who were able to state at least 15 (out of 20, above chance at $p<.05$) rules for the use of the target feature were classified as more-aware. The debriefing sessions were recorded and transcribed in order to provide a record of whether the learners were able to state correctly the critical number of rules and/or examples.

Analyses and Hypotheses

Answers to the first three research questions, which concern learning in the two instruction groups, were investigated by testing the following three hypotheses:

> H1: Learners in both groups will have higher correct scores on the posttest than on the pretest.
>
> H2: There will be a significantly greater decrease in correct scores on the posttest compared to the pretest by learners in the explicit instruction group.
>
> H3: Learners in the explicit instruction group will make fewer errors on novel items on the posttest than learners in the incidental instruction group.

The answers to research questions four through seven, which concern learners in the more-aware and less-aware learner groups, were investigated by testing the following 4 hypotheses:

> H4: A significantly greater number of learners in the explicit instruction group than in the incidental instruction group will fall into the more-aware learner group.
>
> H5: There will be a significantly greater increase in correct scores on the posttest compared to the pretest by learners in the more-aware learner group.
>
> H6: Learners in the more-aware learner group will make fewer errors on novel items on the posttest than learners in the less-aware learner group.
>
> H7: Learners in the more-aware learner group will continue to show greater learning than learners in the less-aware learner group on the delayed posttest.

The final research question motivates the following hypothesis:

> H8: Average response times by all learners will be faster on previously encountered items than on novel items on the posttest.

Two repeated-measures MANOVAs and follow-up univariate ANOVAs were used to test H1, H2, H3, H5, H6, and H8. Eta-squared was used to test the strength of association of all significant results. The relationship between awareness and performance on the posttest was also tested through a correlational analysis of the number of correct rules learners are able to provide during the debriefing interview and the number correct on the posttest. A pair of two-by-two chi-square tests were planned to test H4 and H7, but as described below, H7 was not tested. Because of the large number of statistical tests, alpha was set at the rather conservative level of .01.

The first MANOVA was used to test for differences between instruction groups. The independent variables for this MANOVA were instruction group (explicit or incidental), and item type (repeated or novel). The dependent variables were correct score (repeated measure, pretest and posttest) and response time (repeated measure, pretest and posttest).

Because the grouping of learners in the learner groups was different from their grouping in the instruction groups, a second MANOVA was necessary to test for differences between the two learner groups. The independent variables for this MANOVA were learner group (more-aware or less-aware) and item type. The dependent variables were, again, correct score (repeated measure, pretest and posttest) and response time (repeated measure, pretest and posttest).

The first chi-square test was used to compare the number of explicit instruction group learners falling into the more-aware learner group, the number of explicit instruction group learners falling into the less-aware learner group, the number of incidental instruction group learners falling into the more-aware learner group, and the number of incidental instruction group learners falling into the less-aware learner group. There were two variables: instruction group and learning group.

The second chi-square test was to compare the frequencies of correct and incorrect use of the target items by more-aware and less-aware learners on the delayed posttest. The two variables were to be learner group and correctness of use. Again, this test was not actually used, as will be explained in the results section below.

Results

Descriptive statistics of pretest and posttest. One purpose of the pretest was to control for prior learning. Therefore, it was decided to remove learners who scored above chance, 52 or higher ($p<.01$). Eleven of the original 60 learners, six in the explicit instruction condition and five in the incidental instruction condition, were thus removed and replaced. In addition, one of the replacement learners in the explicit instruction condition was also removed and replaced. However, even though prior learning was at least partially controlled for, the pretest results indicate that it still played a role. The mean on the pretest was 44.5. Both the median and the mode were 45. The standard deviation was 4.13. The low and high scores were 28 and 51. The distribution of pretest scores is displayed graphically in Figure 3.1. Learners as a whole seemed to be operating somewhat above chance.

Item facility (IF) scores, that is the proportion of learners who answered an item correctly, were calculated to see which items were already known by the majority of learners. The mean IF was 0.5556, with a standard deviation of 0.1836, a low of 0.2000, and a high of 0.9500. The items with the eight highest and four lowest IFs are displayed in Table 3.1. The four items that were most often answered correctly are names of countries without *the*. Interestingly, none of the items with the highest IFs take *the*, while

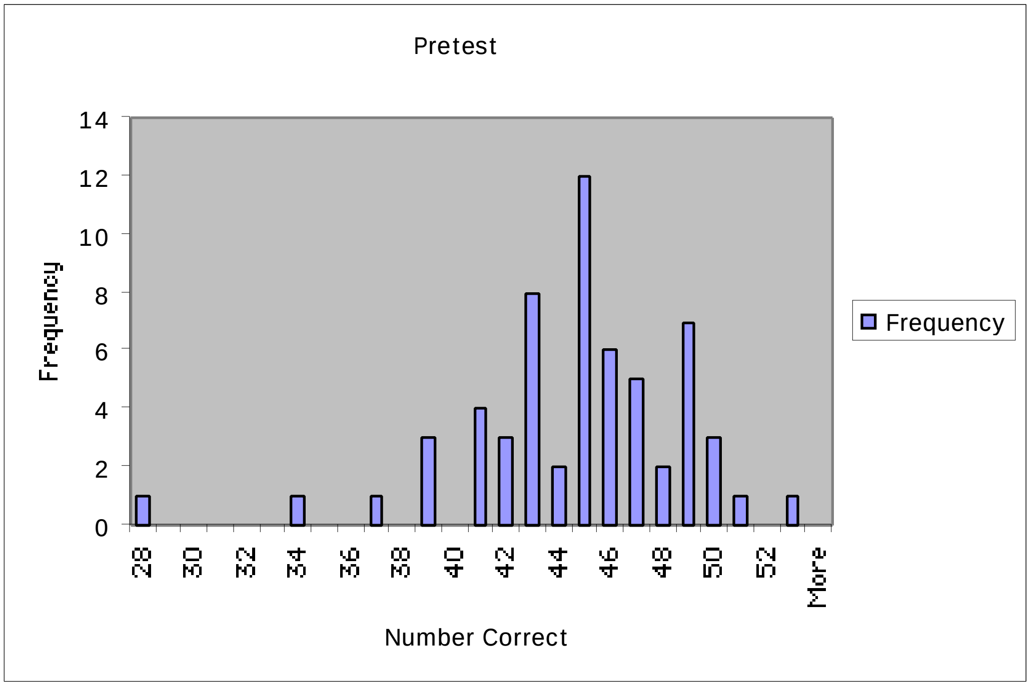

Figure 3.1: Pretest Results

all the items with the lowest IFs do take *the*. This implies that items which do not take *the* were easier than items which do. In order to see whether this is true, all 80 items were placed in a rank order scale from one to 80, with one being the item with the highest IF, that is the easiest item, and 80 being the item with the lowest IF, or the most difficult item. A Mann Whitney U test, which actually had not been previously planned, was then calculated in order to see whether items which take the zero article ranked higher than items which take *the*. The observed *z*-score was -6.086, significant at $p<.01$. Items which take the zero article were indeed easier. This indicates a bias among learners to choose the zero article on the pretest. (Mann Whitney U, a nonparametric test, was used instead of a more powerful parametric test since the relative ease or difficulty of test items was not a central concern of the study and the researcher wanted to avoid problems with probability levels which could be caused by using too many powerful tests.)

Table 3.1: Items with High and Low IF

High IFs	Item	Low IFs	Item
0.95	Canada	0.20	the Falkland Islands
0.95	France	0.27	the Moldive Islands
0.95	Russia	0.28	the Northwest
0.95	Italy	0.30	the Izu Peninsula
0.90	Saipan		
0.90	Tahiti		
0.85	Molokai		
0.82	Kaimuki Elementary School		

Learners were much more spread out on the posttest. The mean was 52.23, with a standard deviation of 12.35, a low of 30, and a high of 78. This greater dispersion was to be expected if some learners learned more than others. The distribution of posttest scores is displayed graphically in Figure 3.2. On the posttest, learners' scores appeared to fall into two groups. In addition, the bias for choosing the zero article over the definite article ded not seem to be playing a role on the posttest. As for the pretest, a Mann Whitney U test was calculated. The *z*-score turned out to be -1.030, not significant at p<.01.

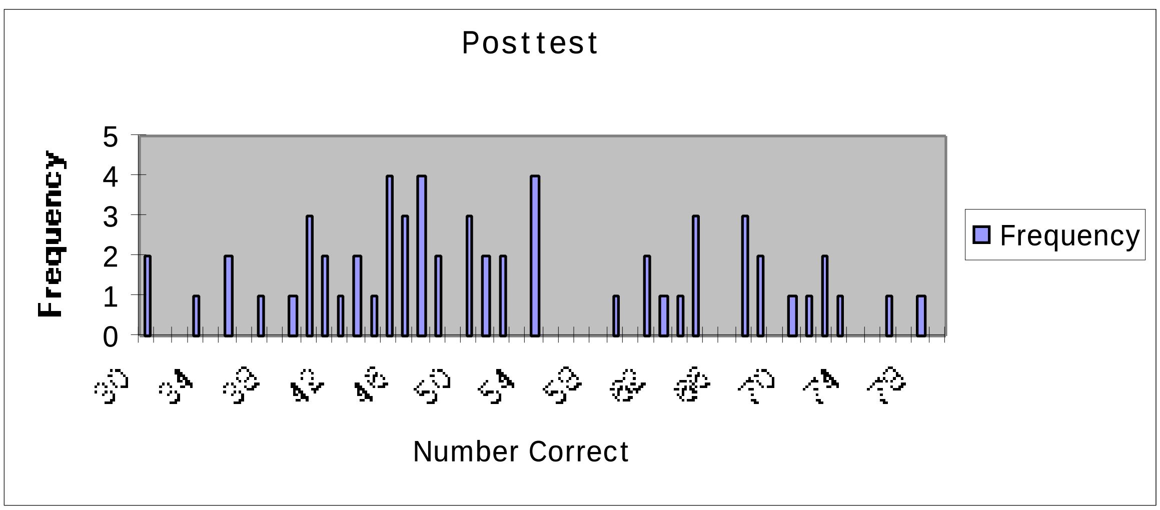

Figure 3.2: Posttest Results

There was a large difference between the standard deviations, and thus the variances of the pretest and the posttest, indicating that the assumption for ANOVA of equivalent variances (Brown, 1988; Hatch & Lazaraton, 1991; Guilford & Fruchter, 1973) may not have been met. The standard deviations on the pretest were, again, 4.13 and 12.35, respectively. This makes the variances of the pretest and posttest 17.0569 and 152.5225, quite a large difference. Applying the F_{max} test of dividing the variance of the posttest by that of the pretest yields an F-ratio of 8.942, which is significant at $p<.01$ (df=58, 1). According to Brown (1988), this should not be a major problem unless there is at least a ratio of three to one between sample sizes. According to Guilford and Fruchter (1973), not meeting this assumption means that the probabilities may be a bit higher than they seem. This means that, with an alpha level of .01, a probability of .008 or .009 might be suspect. While the comparisons between more-aware and less-aware groups described below are unbalanced, the ratio between groups is smaller than three to one. In addition, the probabilities of the significant MANOVA and ANOVA results reported below were .0001, so it is unlikely that the observed difference in variances is a problem.

Reliability estimates were calculated for both the pretest and the posttest. As in the pilot study described above, it was expected that the K-R 20 reliability estimates for the pretest would be extremely low, possibly even negative, but would be much better for the posttest. Indeed, the K-R20 reliability estimate for the pretest was -0.0035 and the correlation between the first 40 and second 40 items was 0.1208, with a full-test reliability, calculated by the Spearman-Brown Prophecy formula, of 0.2156. These extremely low scores probably reflect the large amount of guessing on the pretest. However, the K-R20 reliability estimate for the posttest was 0.9006 and the correlation between the first and second 40 items was 0.9158, with a full-test reliability of 0.9560.

These estimates indicate that, when learners were not relying primarily on guessing, the test could reliably separate learners on the basis of their knowledge of the use of *the* and zero article with place names.

Comparison of instruction groups. Time on task was measured to the minute and was roughly equivalent for both instruction groups. The median time of the incidental instruction group was 21.5 minutes, with a minimum of 11 minutes and a maximum of 32. The median time of the explicit instruction group was 22 minutes, with a minimum of 15 and a maximum of 31.

A repeated-measures MANOVA comparison of instruction groups indicated that there were differences between pretest and posttest correct scores (Wilks' Lambda=0.618, p=0.0001), between groups on their correct scores (Wilks' Lambda=0.560, p=0.0001), and between average response times on the pretest and posttest (Wilks' Lambda=0.400, p=0.0001). Significant effects were not found for any interaction between number correct and item type (repeated versus novel) (p=0.7943) or for any interaction between number correct, item type, and group (p=0.8232). Nor was a significant interaction found between response time and group (p=0.3734), response time and item type (p=0.1862), or response time, item type, and group (p=0.8388).

The pretest and posttest mean scores for each group and the means of each participant's mean response times are shown in Table 3.2. The results of follow-up repeated measures ANOVAs are shown in Tables 3.3 and 3.4. Finally, Table 3.5 gives the strengths of association, eta squared for repeated measures, of the significant ANOVA effects.

Table 3.2: Mean Scores and Response Times

Test	Group	Mean Score	*sd*	Mean RT	*sd*
Pretest	Explicit	44.433	3.657	4442.0597	1313.3685
Pretest	Incidental	44.467	4.703	5930.8375	1203.4157
Posttest	Explicit	60.967	10.270	3020.9929	909.2106
Posttest	Incidental	43.500	6.842	2961.2221	1216.1025

Eighty is the highest possible score. Response times are averaged across items and then across participants; given in milliseconds.

Table 3.3: ANOVA Comparison of Scores and Groups

Source	*df*	Sum of Sq.	Mean Sq.	*F*-value	*p*
Group	1	2279.408	2279.408	40.183	.0001*
Subject	58	3290.083	56.726		
Test	1	1817.408	1817.408	49.460	.0001*
Test * Group	1	2296.875	2296.875	62.508	.0001*
Test * Subject	58	2131.217	36.745		

dependant variable: score; *significant at p<.01

Table 3.4: ANOVA Comparison of Response Times

Source	*df*	Sum of Sq.	Mean Sq.	*F*-value	*p*
Group	1	674,300	674,300	.308	.5811
Subject	58	127,000,000	2,190,000		
Test	1	53,140,000	53,140,000	96.790	.0001*
Test * Group	1	243,800	243,800	.444	.5078
Test * Subject	58	31,840,000	549,000		

dependant variable: response time; *significant at p<.01

Table 3.5: Eta Squared for Significant Effects

Effect	SS Effect	SS Total	Eta Square
Group (correct)	2279.408	11814.991	.193
Test (correct)	1817.408	11814.991	.154
Test * Group (correct)	2296.875	11814.991	.194
Test (response time)	53,140,000	213,000,000	.249

Eta Squared equals SS Effect divided by SS Total

As can be seen in Table 3.2, the explicit instruction group mean score and the incidental instruction group mean score were almost the same on the pretest. However, the posttest mean of the explicit group was much higher than either that group's pretest mean or the posttest mean of the incidental group. In fact, the posttest mean of the incidental group was slightly lower than that group's pretest mean. The significant test by group interaction effect indicates that the difference between the posttest score of the explicit group and the posttest score of the incidental group is significant, as is the difference between the posttest score of the explicit group and that group's pretest score. As can be seen in Table 3.5, the test by group interaction accounts for 19.4 percent of the variance in score data.

As can also be seen in Table 3.2, mean response times dropped for both groups between pretest and posttest. This change is significant and accounts for 24.9% of the variance in the response time data. While the average response time of the incidental group dropped more than that of the explicit group, there was not a significant interaction between group and response time.

Comparison of Learner Groups. Learners were classified as more-aware if they are able to answer correctly at least 15 of the 20 questions on the debriefing questionnaire. Each of these questions corresponds to exactly one rule of thumb for the use of definite and zero articles with place names. Nineteen of the 30 learners in the explicit instruction group were thus classified as more-aware and 11 as less-aware. All 30 learners in the incidental instruction group were classified as less-aware. This difference in frequencies is significant (chi square = 27.8, p<.01, df=1), which is not surprising given that almost two thirds of one group and none of the other group are classified as more-aware. (Chi square was used to compare the frequencies of more-aware and less-aware learners from each instruction group since the study relies on dividing learners into these two learner groups.) Table 3.6 displays the number of questions answered correctly by participants in each group. No participants answered fewer than seven questions correctly.

Table 3.7: Frequency of Participants Correctly Answering a Given Number of Awareness Questions

Correct Answers	Explicit Instruction	Incidental Instruction
20	1	0
19	1	0
18	3	0
17	5	0
16	5	0
15	4	0
14	3	5
13	2	4
12	1	12
11	1	1
10	2	4
9	0	3
8	1	1
7	1	0

Time on task was similar for both learner groups. The median time on task of the less-aware group was 22 minutes, with a minimum of 11 and a maximum of 32. The median time of the more-aware group was also 22, with a minimum of 15 and a maximum of 30.

A repeated-measures MANOVA comparison of learner groups indicates that there were differences between pretest and posttest scores (Wilks' Lambda=0.656, p=0.0001), between groups on the tests (Wilks' Lambda=0.871, p=0.0001), and between average response times on the pretest and posttest (Wilks' Lambda=0.457, p=0.0001). Significant effects were not found for any interaction between test and item type (repeated versus novel) (p=0.9591) or for any interaction between test, item type, and group (p=0.7535). Nor was a significant interaction found between response time and group (p=0.0700), response time and item type (p=0.1605), or response time, item type, and group (p=0.6609). Obviously, these results closely reflect the results of comparisons between instruction groups, with similar differences found between the explicit and incidental instruction groups and the more-aware and less-aware learner groups.

The pretest and posttest mean scores for each learner group and the means of each learner's mean response time are shown in Table 3.7. The results of follow-up repeated-

Table 3.7: Mean Scores and Response Times of Learner Groups

Test	Group	Mean Score	*sd*	Mean RT	*sd*
Pretest	More-aware	45.316	3.742	4466.229	1420.6312
Pretest	Less-aware	44.415	4.387	4255.191	1183.0058
Posttest	More-aware	66.316	7.032	3016.662	947.0381
Posttest	Less-aware	45.707	8.044	2979.265	1126.4222

Eighty is the highest possible score. Response times are averaged across items and then across participants; given in milliseconds.

Table 3.8: Repeated-Measures ANOVA Comparison of Scores and Learner Groups

Source	*df*	Sum of Sq.	Mean Sq.	*F*-value	*p*
Group	1	3003.461	3003.461	66.759	.0001*
Subject	58	2609.406	44.990		
Test	1	3226.123	3226.123	98.938	.0001*
Test * Group	1	2521.223	2521.223	77.320	.0001*
Test * Subject	58	1891.244	32.608		

dependent variable: score; *significant at $p<.01$

Table 3.9: Repeated-Measures ANOVA Comparison of Response Times and Learner Groups

Source	*df*	Sum of Sq.	Mean Sq.	*F*-value	*p*
Group	1	400,700	400,700	.183	.6708
Subject	58	127,300,000	2,195,000		
Test	1	48,220,000	48,220,000	87.700	.0001*
Test * Group	1	195,700	195,700	.356	.5531
Test * Subject	58	31,890,000	549,900		

dependent variable: response time; *significant at $p<.01$

Table 3.10: Eta Squared for Significant Effects

Effect	SS Effect	SS Total	Eta Square
Group (score)	3003.461	13251.457	.227
Test (score)	3226.123	13251.457	.243
Test * Group (score)	2521.223	13251.457	.190
Test (response time)	48,220,000	208,010,000	.232

Eta Squared equals SS Effect divided by SS Total measures ANOVAs are shown in Tables 3.8 and 3.9. Finally, Table 3.10 gives the strengths of association, eta square for repeated measures, of significant ANOVA effects.

As can be seen in Table 3.7, the more-aware group mean score and the less-aware group mean score were almost the same on the pretest. However, the posttest mean of the more-aware group was much higher than either that group's pretest mean or the posttest mean of the less-aware group. The posttest mean of the less-aware group was almost unchanged from its pretest mean. The significant effect for group and the significant interaction between group and test indicate that the difference between the posttest score of the more-aware group and the posttest score of the less-aware group is significant, as is the difference between the posttest score of the more-aware group and that group's pretest score. As can be seen in Table 3.10, the group effect accounts for 22.7% of the variance in the score data and the test by group interaction accounts for 19.0%.

As can also be seen in Table 3.7, response time dropped for both groups between pretest and posttest. This change is significant and accounts for 23.2% of the variance in the response time data. Once again, the results of the learner group comparisons closely reflect the results of the instruction group comparisons.

In dividing learners into more-aware and less-aware groups, a continuous scale, the number of correct answers out of 20, on the debriefing questionnaire, was converted to a two-level nominal scale. In making this conversion, some information about learners' awareness was lost, as can be seen in the fact that five learners in the incidental instruction group missed being classified as more-aware by just one question. In order to compensate for this loss, a correlation was calculated between learners' awareness levels (zero through 20) and their posttest scores. This correlation was 0.7971 (significant at $p<.01$, $r^2=0.6354$). This is surprisingly high, so a correlation between awareness levels and pretest scores was also calculated. This correlation was 0.1988 (not significant at $p<.01$, $r^2=0.0395$). This supports the results of the comparisons between learner groups reported above, indicating a close relationship between awareness and learning.

Delayed Posttest. The delayed posttest was administered to a subset of the original 60 who had participated in the first experiment. It included five learners from the incidental instruction group, all of whom were, of course, classified as less-aware, and five learners from the explicit instruction group, two less-aware and three more-aware. By asking questions such as *What is the highest mountain in the world?* and *Can you name any hotels in Waikiki?*, the researcher was able to elicit a total of 105 tokens of place names which take the zero article and 70 place names which take the definite article. A complete list of the questions that were used in the delayed posttest is provided in Appendix C.

There was an extremely strong bias among these 10 learners to use the zero article with all place names. There were only three instances of the definite article being used incorrectly with place names which take the zero article, while there were only five instances of the definite article being used correctly with place names which take it. The zero article was used correctly 102 times and the definite article was used incorrectly 65 times. Because this bias in favor of no article was so obvious, it was decided that the chi-square test originally planned was not necessary. However, looking at who used the definite article correctly and who used it incorrectly did reveal some interesting information.

The definite article was incorrectly used once each by three different learners. All three learners were in the explicit instruction group. Two of them had been classified as less-aware and one as more-aware. One less-aware learner incorrectly used *the* with the name of a bay, saying*the Hanauma Bay*. This learner had an awareness level of fourteen, meaning that the learner just barely failed to be classified as more-aware. This learner's posttest score was 64. This learner misused the definite article on the delayed posttest even though the question about whether or not to use *the* with bays was correctly answered on the awareness assessment. Another less-aware learner incorrectly used *the* with a part of a continent, saying *the Eastsouth Asia* instead of *Southeast Asia*. This learner had an awareness level of eight and a posttest score of 46. This learner had incorrectly answered the question about whether or not to use *the* with parts of continents. One more-aware learner also incorrectly used *the* with part of a continent, saying *the Middle Asia* instead of *Southeast Asia*. This learner's awareness level was 15, the minimum to be classified as more-aware, with a posttest score of 70. This learner had also incorrectly answered the relevant question on the awareness assessment.

The definite article was used correctly once by one more-aware learner in the explicit instruction group and four times by one less-aware learner in the incidental instruction group. The explicit group, more-aware learner used it correctly with a country name, saying *the United States of America*. This learner had an awareness level of 17 and a posttest score of 54. The relevant question had been answered correctly on the awareness assessment.

The most interesting delayed posttest data comes from the less-aware, incidental group learner who used *the* correctly four times. This learner also used the zero article correctly 10 times and failed to use *the* three times. This learner's awareness level was 13. This learner had a posttest score of 54, the highest in the incidental group. This learner used *the* correctly in the following contexts: once with the name of a country (*the United States*), once with a region (*the East Coast*), and twice with the names of hotels (*the Hilton Hawaiian Village, the Hyatt Regency*). This learner failed to use *the* in the following contexts: once with the name of a university, once with the name of a region, and once with the name of a peninsula. On the awareness assessment, this learner had correctly answered the questions about using *the* with certain country names, with hotels, and with regions. While this learner had also answered the question about peninsulas correctly, the question about using *the* with names of universities containing *of* had been answered incorrectly. Even though it is not possible to say for sure, this learner seems to be displaying some long-term learning. These two learners are further discussed in Chapter Five. (See Appendix D for the transcript of the second learner's delayed posttest.)

In the researcher's opinion, these data about correctly and incorrectly using *the* with place names are interesting, but they are also too meager to be used to make any inferences about long-term effects of the instruction. All that really can be inferred from the posttest data is that learners display little or no long-term learning.

Discussion of Experiment One

The eight hypotheses of Experiment One, described above in the Analyses and Hypotheses section, are repeated here:

H1: Learners in both groups will have higher correct scores on the posttest than on the pretest.

H2: There will be a significantly greater increase in correct scores on the posttest compared to the pretest by learners in the explicit instruction group.

H3: Learners in the explicit instruction group will make fewer errors on novel items on the posttest than learners in the incidental instruction group.

H4: A significantly greater number of learners in the explicit instruction group than in the incidental instruction group will fall into the more-aware learner group.

H5: There will be a significantly greater decrease in correct scores on the posttest compared to the pretest by learners in the more-aware learner group.

H6: Learners in the more-aware learner group will make fewer errors on novel items on the posttest than learners in the less-aware learner group.

H7: Learners in the more-aware learner group will continue to show greater learning than learners in the less-aware learner group on the delayed posttest.

H8: Average response times by all learners will be faster on previously encountered items than on novel items on the posttest.

Hypothesis two was supported. There was a significantly greater increase in correct scores by learners in the explicit instruction group. However, since there was no significant increase by learners in the incidental instruction group, hypothesis one was not supported. Hypothesis three was, in a sense, supported, since learners in the explicit instruction group outperformed learners in the incidental instruction group on novel items. However, since there was no difference between scores on repeated and novel items for either instruction group, this hypothesis seems, in fact, to have been rendered irrelevant. Learners in the explicit instruction group improved on both types of items, novel and repeated, while learners in the incidental instruction group did not improve on either type.

Hypothesis four was supported. A significantly greater number of learners in the explicit instruction group were placed into the more-aware learner group. In fact, absolutely no learners in the incidental instruction group were placed into the more-aware learner group. Hypothesis five was also supported. Learners in the more-aware learner group had significantly better posttest scores. Learners in the less-aware learner group did not. Again, since there was no difference in scores for repeated and novel items, with more-aware learners improving on both and less-aware learners improving on neither, hypothesis six seems to have been rendered irrelevant.

Hypothesis seven was not supported. There was little or no long-term learning found on the delayed posttest by either more-aware or less-aware learners.

Hypothesis eight was not supported. Average response times dropped for both instruction groups and for both awareness groups between the pretest and the posttest. This is possibly due to a desire on the part of the learners to finish the experiment and collect their five dollars. Unlike what was found in Robinson (1995a, 1996a, 1996b, 1997a, 1997b), learners displaying less learning did not have significantly lower average response times, which would have indicated that they were guessing much more than learners who had learned more.

The implications and limitations of this experiment and of the second experiment are discussed in Chapter Five.

Chapter Four
Experiment Two

Introduction

In Experiment One, the amount of input, as measured by number of examples, was equivalent for both instruction groups. Each group was exposed to forty examples of the target feature, with two examples for each rule of thumb. The explicit instruction group saw each rule together with one example. They then saw twenty more examples, one for each rule. The incidental instruction group saw each example one time, embedded in a statement to which they were instructed to decide whether it was true or false. Time on task was also roughly equivalent. However, for learners who attempt to remember the rules for the use of articles with place names, the incidental instruction presents a much more difficult challenge since learners must extract the rules from the input as well as remember them. Learners receiving explicit instruction, though, needed only try to remember the rules.

The result was that no learners in the incidental instruction group were classified as more-aware, though several came close. In order to compare more-aware and less-aware learners who have received incidental instruction, it was decided to conduct a second experiment in which all learners would receive incidental instruction containing a greater amount of input than in Experiment One. In addition, it was thought that a second experiment could present learners with multiple exposure to the same examples, which may result in a response time effect for repeated items which was not found in the first experiment. By increasing the amount of input, the researcher hoped that the more-aware/less-aware split observed in the explicit instruction group would also be observed in a group receiving only incidental instruction.

Because Experiment Two contains only an incidental instruction group, it does not address research questions one through four, which deal with comparisons between explicit and incidental instruction. In addition, Experiment Two does not contain a delayed posttest, so it does not address research question seven either. Experiment Two does address research questions five, six, and eight, which are repeated here and renumbered (*e2* stands for Experiment Two):

> 1(e2). Will learners in the more-aware learner group display greater improvement than learners in the less-aware learner group, as predicted by the noticing hypothesis and as found in Robinson (1995a, 1997a) and Leow (1997)?
>
> 2(e2). Will learners in the more-aware learner group be better able to generalize from examples on the pretest and used in instruction to novel items on the posttest, as measured by error rates?
>
> 3(e2). Will there be differences in response times to previously encountered items and novel items?

Of course, an important methodological question also needs to be considered. It is possible that even with increased input, no learners will be classified as more-aware, or that there will be so few more-aware learners that it will be impossible to use powerful statistical tests (MANOVA and ANOVA) to compare the two awareness groups.

A few learners in the incidental instruction group in Experiment One made comments after the experiment about how they responded to the true/false statements. A few mentioned that they had been tricked--they had been instructed to decide whether statements were true or false, but were then tested on which names took *the*. A few others mentioned that they were looking for rules but that it was too difficult to find them. These comments were made after the experiment was over and were not elicited by the researcher. For Experiment Two, the researcher decided that it would be interesting to ask learners whether or not they had been looking for rules on the use of the definite article. If some learners look for rules while others do not, the researcher expects that those who do look for rules are more likely to be classified as more-aware. This leads to a fourth research question:

> 4(e2): Will more learners who look for rules be classified as more-aware than those who do not look for rules?

In Experiment One prior learning was at least partially controlled for by removing learners who scored too high on the pretest. Even so, analysis of the pretest results indicated that prior learning still seemed to be playing a role. Therefore, it was decided to remove the easiest items, country names which take the zero article, from Experiment Two. Even though pilot testing had revealed that country names which take the zero article were very easy, country names were retained in Experiment One since the contrast between country names that take the zero article with those which take the definite article seemed difficult. In Experiment Two, both types of country names were removed and replaced with names of libraries (*Hamilton Library*) and deserts (*the Sahara Desert*). It was thought that these changes would result in better control for prior learning and would have the added benefit of having to remove and replace fewer learners. No other changes were made to the items used on the pretest and the posttest. (See Appendix E for the replacement items.)

Method

Participants. Thirty young adult Japanese learners of English as a second language were recruited from among the same population used in Experiment One. None of the learners had participated in Experiment One. All learners were paid five dollars for their participation. Again, in order to control for prior learning, learners who scored too high on the pretest (52 or higher, above chance at $p<.01$) were removed from the study.

Materials. As in experiment one, all materials were computerized using PsyScope software (Cohen, et. al, 1993). Except for the changed items, the pretest and the practice session used before the pretest were identical to those of Experiment One.

Instruction was similar to the incidental instruction used in Experiment One, except that the amount of input was increased. In Experiment One, 40 examples were presented one time each, while in Experiment Two, each example was presented four times. As in Experiment One, learners were presented with statements, one at a time, containing an example of the target feature. Learners were instructed to respond to the statements by deciding if they were true or false. They then pressed the space bar to learn the correct answer. In addition to the 40 statements used in Experiment One, 40 additional statements were also used, each one containing one of the same examples. After responding to these 80 problems, the learners then answered all 80 true/false questions a second time. Learners thus received four times as much input than in Experiment One, were exposed to each example four times rather than once, and were exposed to an example of each rule of thumb eight times rather than twice.

Unlike in Experiment One, instruction was not followed by a distracter task. It was removed in order to prevent the experiment from taking too long. Except for the changed items, the posttest was also identical to that used in Experiment One. Again, 40 of the items are items which were used on the pretest and in the instruction, while the other 40 were novel items.

Assessing learner awareness. The measure of awareness used in Experiment Two was almost identical to the measure used in Experiment One, except that learners were also asked, at the beginning of the interview, whether they had been trying to learn rules during the instruction. Also, it was now unimportant whether learners were aware of rules of thumb for country names. Instead, they were asked whether they use *the* with the names of libraries and deserts.

In Experiment One, the threshold to be classified as more-aware was 15 correct answers out of 20, above chance at $p<.05$. It was decided that this would be the tentative criterion used in Experiment Two as well. However, it was also decided that this level might be lowered if too few learners reached it. Lowering the threshold to 14 correct answers would be above chance at $p<.10$. Lowering it to 13 correct would be above chance at $p<.25$. (See the Results section for the criterion that was actually adopted.)

Analyses

Answers to the research questions were investigated by testing the following four hypotheses:

> H1(e2): There will be a significantly greater increase in correct scores on the posttest compared to the pretest by learners in the more-aware learner group.

H2(e2): Learners in the more-aware learner group will make fewer errors on novel items on the posttest than learners in the less-aware learning group.

H3(e2): Average response times by all learners will be faster on previously encountered items than on novel items on the posttest.

H4(e2): A greater number of learners who look for rules will be classified as more-aware.

A repeated-measures MANOVA, followed by univariate ANOVAs, was planned to test the first three hypotheses. Eta-squared was to test the strength of association of all significant results. The relationship between awareness and performance on the posttest was also tested through a correlational analysis of the number of correct rules learners were able to provide during the debriefing interview and the number correct on the posttest. The independent variables for the MANOVA were learner group (more-aware or less-aware) and item type (repeated or novel). The dependent variables were correct score (repeated measure, pretest and posttest) and response time (repeated measure, pretest and posttest).

Chi-square was used to test the fourth hypothesis. The variables were whether learners were looking for rules (yes or no) and learning group (more-aware or less-aware).

Results

Descriptive statistics of pretest and posttest. The pretest results were somewhat different from those in Experiment One, indicating that the more difficult test did a better job of controlling for prior learning. Three of the original 30 learners were removed and replaced because they scored too high on the pretest (52 or higher, above chance at

p<.01). One of the replacement learners was also removed and replaced. This means that 34 learners had to be tested in order to gain a usable sample of 30, compared to the 72 learners that had to be tested in Experiment One in order to gain a usable sample of 60. Descriptive statistics of the pretest also indicate that prior learning was better controlled. The mean on the pretest was 40.63 out of 80. The median was 40 and there were two modes, 39 and 43. The standard deviation was 4.59. The low was 30 and the high was 51. The distribution of pretest scores is displayed graphically in Figure 4.1.

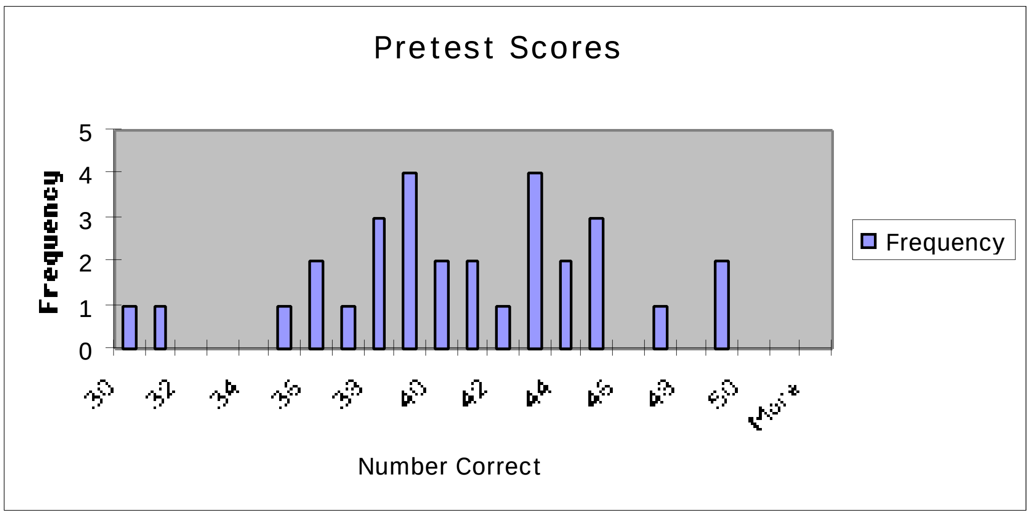

Figure 4.1: Pretest Results

Item facility (IF), that is the proportion of learners who answered an item correctly, was calculated to see which items were already known by the majority of learners. The mean IF was 0.51, with a standard deviation of 0.21, a low of 0.17, and a high of 0.97. The items with the six highest and seven lowest IFs are displayed in Table 4.1. As in the pretest in Experiment One, names of single islands were among the easiest. One of the new items, *the Gobi Desert*, was one of the most difficult. Again as in the pretest in Experiment One, all of the easiest items take the zero article and all of the most difficult items take the definite article. To see whether items which take the zero article were easier, all 80 items were again placed in a rank order scale from one to 80, with one being the easiest item and 80 the most difficult. A Mann Whitney U test was then calculated and the observed *z*-score was -4.206, significant at $p<.01$. There was again a bias among learners to choose the zero article on the pretest.

Table 4.1: Items with High and Low IF

High IFs	Item	Low IFs	Item
0.97	Saipan	0.20	the East Coast
0.97	Tahiti	0.20	the Southeast
0.87	Hanauma Bay	0.20	the Moldive Islands
0.87	Molokai	0.20	the Kahala Mandarin Hotel
0.83	Western Australia	0.17	the Gobi Desert
0.83	San Francisco Bay	0.17	the Northwest
		0.17	the Marshall Islands

The mean on the posttest was 45.90. The median was 43. The mode was 36, but was not a good measure of central tendency on the posttest. The standard deviation was 10.19. The low score was 34 and the high score was 74. The distribution of posttest

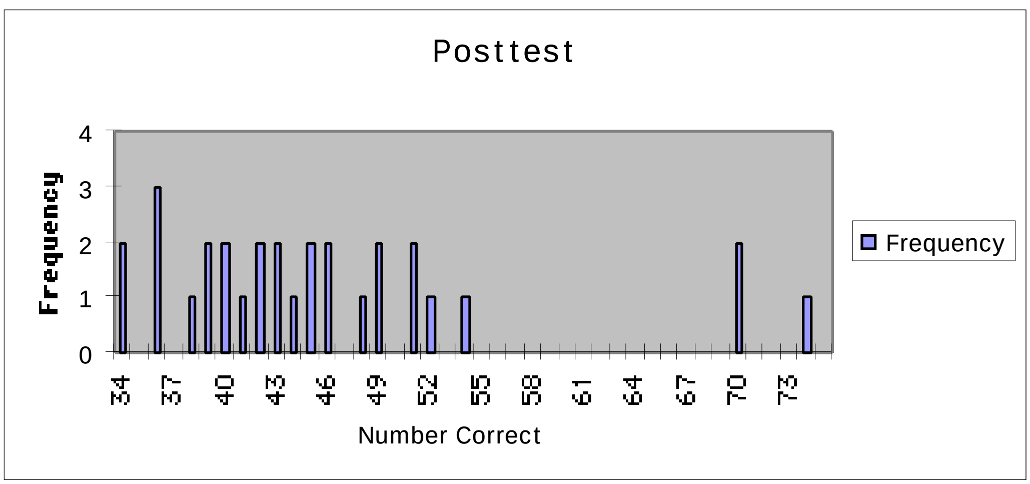

Figure 4.2: Posttest Results

scores is displayed graphically in Figure 4.2. Unlike in Experiment One, there still seems to be a bias for using the zero article. The *z*-score from the Mann Whitney U test was -3.94, significant at $p<.01$.

As in Experiment One, the variances on the pretest and posttest, 21.07 and 103.82 respectively, were quite different. However, the F_{max} ratio, 4.93, was not significant at $p<.01$.

Reliability estimates were calculated for both the pretest and the posttest. These estimates for the pretest were low, but better than for the pretest in Experiment One. The K-R20 reliability estimate for the pretest was 0.2248. The correlation between the first and second 40 items was 0.4215, with a full-test consistency of 0.5930. As in Experiment One, the posttest reliability estimates were much better. The K-R20 reliability estimate was 0.8402. The correlation between the first 40 and second 40 items was 0.8610, with a full-test consistency of 0.9253.

Comparison of learner groups. Only four learners reached the tentative threshold of 15 correct answers to be classified as more-aware. This was not enough learners in the more-aware group to carry out powerful nonparametric tests such as MANOVA and ANOVA (Hatch & Lazaraton, 1991). In order to get at least five learners who could be classified as more-aware, the threshold was lowered to 13 correct answers, above chance at $p<.25$. Under this new criterion, seven learners were classified as more-aware and 23 as less-aware. Since these groups were extremely unbalanced and the threshold to be classified as more-aware allowed a large role for chance, the results of the comparisons between learner groups must, therefore, be interpreted cautiously.

All seven learners who were classified as more-aware stated that they had been looking for rules. However, the majority of learners in the experiment stated that they had been looking for rules. Only five learners stated that they had not. This is interesting in that learners were instructed to respond to meaning, not to look for rules. In a sense, only five out of 30 learners were following directions. The awareness levels of learners who were and were not looking for rules are displayed in Table 4.2. Looking for rules seems to be a precondition for being classified as more-aware, but it is not a very good predictor. The chi-square comparison of learners who were looking for rules and classified as more-aware (*n*=7), were not looking for rules and were classified as more-aware (*n*=0), were looking for rules and classified as less-aware (*n*=18), and were not looking for rules and classified as less-aware (*n*=5) did not reveal any significant differences (chi^2=1.8261, not significant at p<.01). This was probably due to the fact that the majority of learners were looking for rules but most of them were nevertheless classified as less-aware.

Table 4.2: Frequency of Learners Correctly Answering a Given Number of Awareness Questions

Correct Answers	Looking for Rules	Not Looking
20	0	0
19	1	0
18	2	0
17	0	0
16	0	0
15	1	0
14	0	0
13	3	0
12	3	0
11	5	2
10	6	1
9	2	1
8	1	1
7	1	0

Time on task, measured to the minute, was longer than it was in Experiment One, which is not surprising since the learners received four times as much input in Experiment Two. The median time was 30.5 minutes, with a minimum of 20 minutes and a maximum of 51 minutes.

A repeated-measures MANOVA comparison of awareness groups indicated that there were significant differences between pretest and posttest scores (Wilks' Lambda=0.532, p=.0001), a significant interaction between awareness group and test scores (Wilks' Lambda=0.660, p=.0001), and between pretest and posttest response times (Wilks' Lambda=0.469, p=.0001). Significant effects were not found for any interaction between test scores and item type (repeated versus novel) (p=.3947) or for any interaction between test scores, awareness group, and item type (p=.3947). Nor were significant effects found for any interaction between awareness group and response time (p=.5558), between response time and item type (p=.1995), or between response time, awareness group, and item type (p=.6132). These results mirror the comparisons made between learner groups in Experiment One.

The pretest and posttest mean scores and standard deviations and mean response times and standard deviations for each learner group are shown in Table 4.3. The results of follow-up repeated-measures ANOVAs are shown in Tables 4.4 and 4.5. Finally, Table 4.6 gives the strengths of association, eta square for repeated measures, of significant ANOVA effects.

As can be seen in Table 4.3, the more-aware group mean score and the less-aware group mean score were almost the same on the pretest. However, the posttest mean of the more-aware group was much higher while that of the less-aware group was almost unchanged. The ANOVA results indicate that the differences between the posttest scores of the more-aware group and the other test scores were significant. These are the same results as were found in Experiment One. As can be seen in Table 4.6, the group effect accounted for 21.8% of the variance in the score data and the test by group interaction accounted for 11.2%.

Table 4.3: Mean Scores and Response Times of Learner Groups

Test	Group	Mean Score	*sd*	Mean RT	*sd*
Pretest	More	42.714	3.773	5632.361	1458.317
Pretest	Less	40.000	4.700	4652.466	2048.633
Posttest	More	58.429	12.765	3560.800	421.411
Posttest	Less	42.087	5.222	2867.777	1108.835

Eighty is the highest possible score. Latencies are averaged across items and then across learners; given in milliseconds.

Table 4.4: Repeated-Measures ANOVA Comparison of Scores and Learner Groups

Source	*df*	Sum of Sq.	Mean Sq.	*F*-value	*p*
Group	1	974.392	974.392	18.955	.0002*
Subject	28	1439.342	51.405		
Test	1	850.306	850.306	33.551	.0001*
Test * Group	1	498.306	498.306	19.662	.0001*
Test * Subject	28	709.627	25.344		

dependent variable: score; *significant at $p<.01$

Table 4.5: Repeated-Measures ANOVA Comparison of RTs and Learner Groups

Source	*df*	Sum of Sq.	Mean Sq.	*F*-value	*p*
Group	1	7,509,726	7,509,726	2.091	.1592
Subject	28	100,500,000	3,590,641		
Test	1	39,902,945	39,902,945	34.200	.0001*
Test * Group	1	220,826	220,826	.189	.6669
Test * Subject	28	32,668,834	1,166,744		

dependent variable: response time; *significant at $p<.01$

Table 4.6: Eta Squared for Significant Effects

Effect	SS Effect	SS Total	Eta Square
Group (score)	974.392	4471.973	.218
Test (score)	850.306	4471.973	.190
Test * Group (score)	498.306	4471.973	.112
Test (response time)	39,902,945	180,802,331	.221

Eta Squared equals SS Effect divided by SS Total

The only significant change in the response time data was between the pretest and the posttest. As in Experiment One, learners responded to items more quickly on the posttest regardless of learner group.

For the most part, correlations between learners' pretest and posttest scores and awareness levels supported the results of the comparisons reported above. The correlation between posttest scores and awareness levels was 0.8839, significant at $p<.01$. This was a strong correlation, and r^2 was 0.7812. However, the correlation between pretest scores and awareness levels, 0.4335, was also significant at $p<.01$. While this correlation was significant, it was much lower than the correlation between posttest scores and awareness levels. It was also not a very strong correlation, with r^2 of only 0.1879. It indicates, though, that learners who achieved higher awareness levels did have an initial advantage.

Discussion of Experiment Two

The four hypotheses investigated in Experiment Two, reported above, are repeated here:

> H1(e2): There will be a significantly greater increase in correct scores on the posttest compared to the pretest by learners in the more-aware learner group.
>
> H2(e2): Learners in the more-aware learner group will make fewer errors on novel items on the posttest than learners in the less-aware learning group.
>
> H3(e2): Average response times by all learners will be faster on previously encountered items than on novel items on the posttest.
>
> H4(e2): A greater number of learners who look for rules will be classified as more-aware.

Hypothesis one was supported. More-aware learners showed improvement on the posttest while less-aware learners did not. As in Experiment One, there were no effects related to item type, either for correct scores or for response times. Since more-aware learners did better on novel items than less-aware learners, hypothesis two was, in a sense, supported, but the lack of differences between item types seems, again, to have rendered this hypothesis irrelevant. More-aware learners improved on all items regardless of type, while less-aware learners showed improvement on neither type. Hypotheses three was not supported in any sense, a somewhat surprising result given the number of exposures to each repeated item in this experiment. Surprisingly, hypothesis four was also not supported, even though only learners who were looking for rules were classified as more-aware. This is probably because looking for rules was a precondition for becoming aware, but it was not a predictor of becoming aware.

As hoped, increasing the input helped bring about a more-aware/less-aware split among learners receiving only incidental instruction. However, in order to have a large enough more-aware learner group, the awareness threshold had to be lowered. Increasing the amount of input did not seem to have any effect on response times to repeated instances.

The next chapter contains an in-depth discussion of the results of both experiments.

Chapter Five
Discussion

Introduction

Laboratory research on SLA has implications for SLA theory and research and for second and foreign language pedagogy. There are also limits to what such research in general can teach us, as well as specific limitations of any given study. This chapter discusses the results of Experiments One and Two with regard to the relationship between instruction, awareness and learning; incidental versus intentional learning; instance-based learning; long-term learning; and the choice of definite or zero article. Also discussed are the limitations of this study and its implications for theory, research, and pedagogy.

General Discussion of Experiments One and Two

Instruction condition, awareness, and learning. The idea for this study was conceived while the researcher was looking into a variety of effects of instruction studies, such as those reviewed in the second chapter. In some of these studies, it seems to be assumed that learning processes and the accompanying learner awareness are determined solely by the type of instruction that they receive, as if explicit instruction guarantees learner awareness of rules and instruction without an explicit component precludes it. For example, in Winitz (1996), it is assumed that so-called explicit and implicit instruction necessarily result in explicit and implicit learning processes, respectively. In this study, the researcher wished not only to look at learning under different instruction conditions, but also to look at the development of awareness under the different conditions and the relationship between awareness and improved performance.

It was expected that awareness would develop under both the explicit and incidental instruction conditions in this study, with the explicit instruction being more facilitative of awareness and thus resulting in more learners who could be classified as more-aware. It was also expected that there would be variability among the learners regardless of instruction condition, with less-aware and more-aware learners in both instruction groups. Since a certain level of awareness has been argued to be necessary for learning (Schmidt, 1990, 1993, 1994b, 1995), it was expected that learners classified as more-aware would outperform those classified as less-aware. And, since explicit instruction is likely to facilitate awareness, it was expected that the explicit instruction group would outperform the incidental instruction group. To a certain extent, these expectations were fulfilled. The expected split between more-aware and less-aware learners within a single instruction condition was found in the explicit instruction group of Experiment One. About two-thirds of the learners were classified as more-aware and the rest as less-aware. In addition, there was a great deal of variability in the awareness of learners in this group, as measured by the number of correct answers on the debriefing questionnaires (see Table 3.6). Indeed, the one learner who correctly answered the least number of questions on the questionnaire, seven, was in this instruction group. This shows that it cannot be assumed that learners who receive some kind of explicit instruction necessarily develop awareness of the targets of instruction. Instead, explicit instruction may facilitate awareness, but this will vary from learner to learner.

There was also, as expected, variability in awareness among learners in the incidental instruction groups in both experiments. However, for the most part, the expected split between more-aware and less-aware learners did not materialize. In Experiment One, several learners came close to the threshold for being classified as more-aware, but no learners actually reached it (see Table 3.6). The researcher then thought that, if the amount of input were increased, some learners in this group would be able to reach the threshold and a more-aware versus less-aware split would appear as in the explicit instruction group. Therefore, Experiment Two, in which all learners were placed in the incidental instruction condition, was designed, with four times as much exposure as in Experiment One. However, in this experiment also, there was not much of a split, with only four learners reaching the original threshold to be classified as more-aware. Even after the threshold was lowered to 13 correct answers, above chance at $p<.25$, only seven learners were classified as more-aware. Given that the learners in this experiment saw two examples per rule four times each, and that the majority of learners claimed to have been looking for rules on the use of *the*, it is rather surprising that they did not do any better. Some possible reasons that they could not learn any better are discussed below.

As expected, the learners in the explicit instruction group outperformed those in the incidental instruction group. In fact, the incidental group learners in Experiment One showed no improvement at all. This reflects the fact that explicit instruction was much more facilitative of awareness. Also as expected, there was a very close relationship between awareness and learning. Not only did the more-aware learners outperform the less-aware learners in Experiment One, but there were also strong correlations between level of awareness, as measured by the number of correct answers on the debriefing questionnaire, and posttest scores in both experiments. In fact, the strength of the correlation in Experiment One was such that the researcher was prompted to calculate the correlation between pretest scores and awareness levels in order to make sure that prior knowledge was not contaminating the data. Even though this correlation was positive, it was not significant and thus cannot be said not to be the result of chance. Even if a type two error has been made, and there actually is a real correlation between pretest scores and awareness levels, the small size of the correlation means that it cannot be a very important factor. There was, though, a significant positive correlation between pretest scores and awareness levels in Experiment Two, indicating that more-aware learners began the experiment with an advantage. However, in contrast to the correlation between posttest scores and awareness levels, this correlation is not very strong. Of course, in Experiment One, the results of the comparisons of the instruction groups and of the learner groups are reflections of each other, with the more-aware group in fact being the explicit instruction group minus those learners who apparently did not benefit much from the explicit instruction. In addition, if learning is directly related to awareness, then it could be said that the posttest, which measured learning, and the debriefing, which measured awareness of rules of thumb, were measuring essentially the same thing.

As in Leow (1997) and Robinson (1995a, 1996a, 1996b, 1997a), the empirical evidence found in this study supports Schmidt's noticing hypothesis in that there was more learning with more awareness. However, the fact that there was a strong relationship between awareness and learning does not uniquely support the idea that awareness is necessary for SLA. It could be, for instance, that the posttest implicates only metalinguistic knowledge. Truscott (1998a) argues that this is a problem with all empirical studies of noticing and that no studies have shown that noticing is necessary for the acquisition of linguistic competence. It could also be that detection, which does not require awareness but often occurs in conjunction with it, is responsible for the improved performance of the more-aware learners (Tomlin & Villa, 1994). The more-aware learners could be more successful because awareness facilitated detection or detection started cognitive processes which brought about awareness, and thus there is a correlation between more-aware learners and learners who had detected enough to display learning. This study has not been designed to separate empirically detection and awareness, and designing an SLA study which could investigate whether learning can take place with detection but without awareness seems to be a practical impossibility, at least at the present time.

This study also provides some evidence against models of learning without awareness (Krashen, 1981, 1982, 1994; Paradis, 1994), although it does not disprove those models. In Krashen's model, for example, unconscious acquisition occurs when input just beyond a learner's current level (input at *i+1*) is somehow comprehended by the learner *and* the learner is in an affective state that allows this input to enter the learner's interlanguage system (low affective filter). It is impossible to prove that the target feature of this study was at the *i+1* level for the less successful learners or that the less successful learners had low affective filters. However, the use of place names was something which the learners had not yet acquired and the input in the incidental condition seems to have been comprehensible to the learners. It is likely that the target feature was at the *i+1* level for at least some of the learners. In addition, while computer anxiety may have caused some learners to raise their affective filters, the researcher was friendly to all the learners, who seemed eager to help and happy to receive five dollars for less than an hour's work. It is therefore reasonable to expect that the majority of learners did not have high affective filters. Also, since in Experiment One all learners were randomly assigned to instruction groups, it is unlikely that two-thirds of the learners in the explicit instruction group were ready to acquire the target feature and had low affective filters while this was true for no one in the incidental instruction group. It may be that Krashen’s model is unfalsifiable, and in the absence of empirical evidence that SLA can occur without awareness, perhaps the burden of proof should be on those who claim it does (c.f., DeKeyser, 1994).

Incidental versus intentional learning. As explained in Chapters One and Three, the incidental condition is called an incidental *instruction* condition rather than an incidental *learning* condition since learners may choose an intentional strategy. It should not have been difficult for any of the learners to realize that the study was about using *the*--the pretest should have alerted them to this. While in Experiment One learners were not asked whether they had been looking for rules, they were asked this in Experiment Two and the majority answered that they had. In spite of the fact that learners were only instructed to decide whether statements were true or false, and thus to respond to meaning, the majority, 25 out of 30, chose to focus on the use of *the* as well. Even though looking for rules was not predictive of being more-aware, all the learners who were classified as more-aware in Experiment Two were part of this majority. It seems safe to say that any learning which happened in this condition was more intentional than incidental. It is even safe to say that the majority of learners who did not show any improvement were also trying to learn intentionally. Since learning in the incidental instruction condition was not actually incidental, learning in this condition will be discussed in the next subsection.

Instance-based learning. Even though no effect for response time between novel and repeated items was found in the study, instance-based learning, or data-based processing as it is called by Robinson (1995b), does seem to have played a role for at least some learners, something which is especially evident in Experiment Two. As explained above, learning in the incidental instruction condition was most likely intentional rather than incidental. However, since learners were provided only with examples, not with rules, learning in this instruction group was almost certainly instance-based rather than rule-based. (Since the explicit instruction contained both rules and examples, learning may have been rule-based, instance-based, or both.) While no incidental instruction group learners in Experiment One reached the threshold to be classified as more-aware, several of them did come close to reaching that threshold, indicating that they may have been able to become aware of at least some of the rules of thumb based on the data alone. In addition, four learners in Experiment Two reached the original threshold and seven reached the lowered threshold to be classified as more-aware. On the other hand, the incidental instruction provided in both Experiments One and Two seems to have benefited other learners, the majority, not at all.

The lack of effect for response time is probably due to the interaction of a number of factors. First, there may not have been enough exposure to each instance, even in Experiment Two, for direct retrieval of an instance to be noticeably faster than application of a rule. While Logan (1990) argues that automaticity and repetition priming are both based on direct instance retrieval, and there can thus be decreased response time in recognition with even just one prior exposure, this does not seem to have applied in this study, in which learners were asked to make a correct choice rather than to decide, for example, whether a string of letters was a word or nonword. Second, instead of pressing a button as soon as an instance had been retrieved from memory, learners may have continued to think about their choices on the posttest even after retrieving an appropriate instance. Their response times only show how long learners took to answer the items, and thus only indirectly how long it took to retrieve the relevant information from memory. Third, the overall decrease in response times between the pretest and posttest, possibly reflecting learners' desire to finish, would have obscured any very small response time effects that may have occurred. For example, in Logan (1990) decreases in response times after between two and 10 presentations were less than 200 milliseconds.

One more factor may have also worked against faster responses to repeated versus novel items. As described in Chapter One, similarity between repeated items and novel items can effect response times on novel items. Palmeri (1997) found that the degree of similarity between exemplars and novel items was related to response times, with a higher degree of similarity resulting in a quicker response time. Unlike in Palmeri, it is not possible to calculate the degree of similarity between repeated and novel items in this study, such as *the Nile River* (repeated) and *the Amazon River* (novel). However, there are 20 categories of place names in this study, each of which has both repeated and novel items, and it is likely that items in the same category are similar enough for this to affect the time needed to make judgements on the posttest.

That some learners but not others were able to learn something directly from the instances is demonstrated by the results of Experiment Two. Even though all learners were exposed to the forty examples, two for each rule, four times each, only four learners reached the original threshold to be classified as more-aware. These four learners had higher posttest scores than any other learners in the experiment. Since the researcher was unable to make MANOVA comparisons due to the small number of more-aware learners, it was decided to lower the threshold to be classified as more-aware. Even so, only seven learners were so classified. While it was pointed out that the results of Experiment Two need to be treated cautiously, the comparisons of more-aware and less-aware learners indicate that only the few more-aware learners learned anything. In contrast, even with the lowered threshold, twenty-three learners were still unable to gain enough rule awareness to be classified as more-aware, despite the fact that the majority of learners stated that they had been looking for rules on the use of *the*, and these learners did not show measurable improvement on the posttest. Of course, it should be noted again that reaching the original more-aware threshold in Experiment Two actually required more work than it did in Experiment One, since removing country names and replacing them with names of deserts and libraries had the effect of raising that threshold.

As discussed in Chapter One, theoretical proposals bringing instance-based learning into SLA have been made by Robinson (1995b), who calls it data-based processing, and Truscott (1998b). Truscott's proposal involves recasting the acquisition/learning distinction in terms of learning being the development of algorithms and acquisition the accumulation of instances. While Truscott does not actually mention whether awareness is involved, seeing the accumulation of instances in these terms implies that it occurs in the absence of awareness. Robinson, though, argues that awareness plays a role in both data-based processing (instance-based learning) and conceptually-based processing (rule-based learning). The close relationship between successful instance-based learning and awareness found in Experiment Two supports Robinson's position.

It may be surprising that the incidental instruction groups, especially in Experiment Two, did not do any better. According to Krashen (1994), learners can achieve a high level of competence in a second language with only comprehensible input and without receiving explicit instruction. Indeed, that naturalistic second language learners can develop a high level of communicative competence can be seen in the case studies of naturalistic learners conducted by Schmidt (1983) and Klein (1986), which provide empirical evidence that learners can learn directly from the input and achieve a high level of communicative competence without explicit instruction. Through incidental learning, they can, for example, gain the ability to make themselves understood in the target language, to understand other users of the target language, and to behave in pragmatically appropriate ways. However, what seems more difficult is for the learners in these studies to develop target-like usage of the morphology and syntax of the second language directly from the input, even after long periods of contact and interaction. One Spanish migrant to Germany, for instance, is described by Klein as having "no inflection whatever at his disposal," but is nevertheless "a skilled story-teller" (p. 126).

In Experiment Two of this study, the majority of learners claimed to have been looking for rules but either failed to find them or misinterpreted them. While Krashen (1994) claims that the grammar of natural languages is too complex to be explicitly taught, it may in fact be that it is too complex to be learned directly from the input, at least by the majority of learners. Klein (1986) divides the learner's task into four problems, one of which is the problem of analysis, or figuring out the rules, with or without awareness, of the target language. It is possible that when it comes to the formal features of the target language, such as the arbitrary use of the definite article with certain categories of place names in English, this problem of analysis is especially difficult. It may be that only very few learners are capable of figuring out the grammar of a second language on their own and that the majority of learners actually require the facilitative aid of explicit instruction in order to become aware of some formal features.

One Experiment Two learner in particular demonstrates the ability of learners not to learn anything through instance-based learning. Like most learners, this learner claimed to have been looking for rules on the use of *the*. However, on the debriefing questionnaire, this learner answered that all item categories do not take *the*, resulting in an awareness level of ten. A look at this learner's pretest scores reveals that the zero article was chosen for each item. On the posttest, the zero article was chosen for all but two items, for both of which the definite article was chosen incorrectly, making this learner's posttest score, thirty-eight, lower than the pretest score of forty. (This learner took 36 minutes to complete the experiment, above the median time of 30.5 minutes, indicating that a strategy of just pressing buttons to get through the experiment was not employed.) It may be said that this use of the definite article was not at the *i+1* level, so the learner was unable to learn anything, or it may be said that the input was not comprehensible or that the learner's affective filter was high. However, the researcher knows this learner personally. This learner is married to someone who speaks a standard outer circle dialect of English (Kachru & Nelson, 1996) and does not speak Japanese. All communication between spouses is in English. The learner is communicatively fluent and has a high level of comprehension in both oral and written modalities. It is highly unlikely that the input was not comprehensible. Also, it seems to the researcher that this learner is as or more proficient than those learners who did learn something, so it seems strange that this use of articles would be above *i+1*. Finally, the learner does not seem at all to be intimidated by or to dislike either computers or the researcher, so it would also be strange to say that the learner had a high affective filter. Still, the learner apparently learned nothing from the input. According to Klein (1986), another of the four problems faced by learners is the problem of matching, or figuring out how the learner's variety of the language is different from the norms of the target language. This matching problem actually becomes more difficult as the learner becomes more proficient, since the learner's language more closely approximates the target, and the learner is more certain of the accuracy of the language rules he or she has. Perhaps this learner, as well as some of the other unsuccessful learners, was extremely confident that all place names take the zero article, and therefore did not notice any discrepancy between the input and this learner's own rules. In addition, the arbitrary and apparently meaningless nature of this use of *the* may have made it more difficult to notice this discrepancy.

While this study found evidence supporting a role for instance-based learning in SLA and for a relationship between instance-based learning and awareness, it is difficult to explain the differences found among learners. Again, the majority of learners in Experiment Two claimed to have been looking for rules, but did not find them. This failure does not seem to be related to proficiency. It is this researcher's subjective opinion that the successful learners were among the more proficient, but that there were also less successful learners of equal or greater proficiency, such as the one described above. It is also difficult to say exactly what learners, particularly the more successful learners, were actually doing. Were they memorizing examples and then using the examples they had memorized to answer the items on the posttest, either through direct retrieval of or analogy to an example, and only constructing rules when they were asked for them during the debriefing? Or were they actively constructing rules of thumb as they encountered the examples? Or were some learners using one strategy and others the other strategy? Unfortunately, the answers to such questions cannot be found in the results of this study.

Long-term learning. Little evidence of any long-term learning was found among the subset of ten learners who had participated in the first experiment and took the delayed posttest. Only two learners used the definite article correctly on the delayed posttest, and both of these learners also produced tokens which should have had the definite article but did not. In addition, it is difficult to determine whether their performance on the posttest can be tied to something they had learned during the experiment. In fact, the more-aware learner from the explicit instruction group who produced one correct token marked with the definite article, *the United States of America*, had answered all four items of this category correctly on both the pretest and the posttest, including the pretest item, *the United States*. It is possible that the explicit instruction this learner received during the experiment reinforced this learner's control over this use of the definite article, but there is no way to prove this one way or the other.

For the less-aware learner from the incidental instruction group who produced four correct tokens marked with the definite article, the results are somewhat more supportive that some long-learning has taken place. The learner got one of these tokens, *the United States*, correct on the pretest. However, only one other of the four items of this category was answered correctly on the pretest. On the posttest, where *the United States* was not an item, this learner answered three of the four items of this category correctly. Another item that this learner produced correctly on the delayed posttest, *the East Coast*, was answered incorrectly on the pretest but correctly on the posttest. In addition, this learner only answered one item of this category correctly on the pretest, but answered three of four correctly on the posttest. The other two tokens that this learner produced correctly on the delayed posttest were names of hotels, *the Hilton Hawaiian Village* and *the Hyatt Regency*. On the pretest, this learner correctly answered one of the four items of this category. The correctly answered item was *the Hilton Hotel*. This learner correctly answered three of the four items of this correctly on the posttest, including *the Hilton Hotel*. (*The Hyatt Regency* was not an item on either the pretest or posttest.) It must be remembered that, while not classified as more-aware, this learner had a higher level of awareness than most other learners in the incidental instruction group of Experiment One. Again, though, it cannot be proven that this learner's delayed posttest performance is tied to learning which took place during the experiment, but this is certainly a more likely possibility than for the other learner.

It is rather disappointing, but hardly surprising, that there was not more evidence of long-term learning on the delayed posttest. One short instruction session cannot be expected to have a very strong long-term influence on learners' productive abilities. In order for long-term learning actually to have taken place, learners would probably have had to receive further relevant input during the period between the experiment and the delayed posttest. In addition, since the delayed posttest required learners to produce rather than just choose the correct form, learners would also probably have needed some relevant production practice during this interval. The researcher has no way of knowing whether the learners had relevant input or practice, but it seems likely that they did not have much.

Zero article or no article? One rather implausible way of viewing the English article system is that it consists of three articles: the indefinite article (*a*/*an*), the definite article (*the*), and the zero article (no surface form). Even though under most conditions the choice would not be conscious, one of these three articles would have to be chosen every time a noun phrase were produced. The indefinite article would be chosen when the noun phrase did not contain any other determiner and referred to something unspecified and singular. The definite article would be chosen when the noun phrase did not contain any other determiner and referred to something either specified in the discourse or specific by context or by nature. The definite article would also be chosen with certain proper nouns, including certain place names, or when specifying one of several identical proper nouns (e.g., *the Eric who is married to Mariko*). The zero article would be chosen when there were another determiner, when the noun phrase referred to something not specified and either plural or mass, and with certain proper names, again including certain place names. (Of course, in most cases, rather than choosing the definite or zero article each time they produced a proper noun, fluent speakers would just retrieve the proper noun with the correct article attached directly from memory.)

However, speakers of English who did not have a target-like article system, a group which includes the majority of Japanese speakers of English as a second language, would probably not make this choice when producing noun phrases in English. The rank-order statistical tests of the ease of the test items described in the last two chapters indicate that items which take the zero article were easier on the pretests in both experiments and on the posttest in Experiment Two. In addition, on the delayed posttest in Experiment One, tokens which take the zero article were rarely marked incorrectly with the definite article, while tokens which take the definite article were rarely marked correctly. This does not indicate, though, a preference on the part of learners for choosing the zero article over the definite article. Rather, the learners were most likely not making any choice between the three articles. Since a noun phrase with no article, if it were target-like in all other respects, would be identical to a noun phrase marked with the zero article, learners who did not make a choice of article would appear to display a preference for choosing the zero article.

At the beginning of this subsection, it was noted that this way of viewing the English article system is not very plausible. The reason is that it is extremely unlikely that noun phrases without either the indefinite or definite article are actually marked with a zero article which does not have any surface form. This implausible scenario was described above in order to illustrate that what most learners were probably doing was not incorrectly not choosing the definite article, but, rather, not making any choice at all.

It would be interesting to see whether there are learners with a fairly target-like article system who nevertheless do not use the definite article with place names which take it, and whether such learners would be better able to learn this use of articles directly from input, without the facilitative aid of explicit instruction. Unfortunately, though, even if such learners exist finding them would be extremely difficult.

Limitations

The major limitation of this study is its low external validity. Laboratory studies of SLA, such as this one, can achieve a higher degree of control than can classroom studies or studies of naturalistic language learners over such variables as prior learning and language exposure outside the classroom or laboratory, but such control is achieved at the sacrifice of some generalizability to non-laboratory contexts. Both the explicit instruction and the incidental instruction used in this study are rather different than instruction which is likely to take place in foreign or second language classrooms. This difference limits the extent to which the findings of this study can generalize to language classrooms. The input in the incidental instruction condition is also different from input likely to be encountered during naturalistic language learning, again limiting its generalizability.

Though it does not necessarily happen in the absence of awareness, incidental learning takes place when learners intend to communicate and are focused on meaning. When, or if, it takes place in language classrooms, it occurs when learners are interacting in the target language, with other learners, teachers, or texts. For example, incidental acquisition of vocabulary seems likely to take place when learners are trying to understand the meaning of a written text. In this study, incidental instruction is operationalized by asking learners to respond to statements which are either true or false, with each statement containing exactly one example of the learning target. It is quite easy for learners to adopt an intentional learning strategy, even though it is much more difficult to use this strategy successfully. It seems unlikely that the incidental instruction in this study is similar to the type of interaction that normally takes place in language classrooms.

Incidental learning also occurs in naturalistic language learning. In classrooms, materials may be developed to promote incidental learning, but this does not happen in naturalistic learning, where learners have to learn what they can from environmental input while trying to interact. Incidental instruction as operationalized in this study, with examples neatly packed into the input, is thus even more different from naturalistic incidental learning than from any incidental learning which is likely to occur in classrooms.

It may seem at first that the explicit instruction is fairly similar to the kind of explicit instruction which is likely to take place in language classrooms. However, there are also important differences between the way explicit instruction is operationalized in this study and the way it occurs in classrooms. In this study, explicit instruction is operationalized as the provision of 20 rules of thumb and two examples of each rule. Learners do not practice applying the rules themselves, except on the posttest. Explicit classroom instruction also involves the provision of pedagogical rules and, most likely, examples as well. However, it seems unlikely that explicit classroom instruction would not involve practice applying the rules. In addition, in communicatively-oriented classrooms, any explicit instruction would also involve communicative interaction with other language users or texts. Such interaction is not included in this study.

Even though differences between the explicit and incidental instruction as operationalized in this study and the instruction and learning which is likely to take place outside the laboratory limit the extent to which the findings of this study can be generalized to other situations, this does not mean that this study, as well as other laboratory studies, teaches us nothing about how languages are actually learned. In particular, since acquisition in this study is superior for more-aware learners, this indicates that awareness is important in SLA and provides some evidence that it plays a causal role. Since awareness is important for acquisition in this study, it is likely that it is also important in SLA occurring in other contexts as well.

Another possible limitation is that the instruction in this study, either explicit or incidental, could be argued to take place over a much shorter time span than instruction in a classroom. However, while learners generally took less time to complete the experiment than the fifty to sixty minutes of instruction typical in a language classroom, they were on task for the duration of the experiment. This is in contrast to what often happens in the language classroom, where learners are typically on task for only a fraction of the class time.

Another limitation of this study is that it does not look at the development of fluency in a second language. Second language acquisition is not just the acquisition of features of the second language. It is also the development of the ability to produce and comprehend language containing those features automatically, without having to devote a great deal of attentional resources. Fluency can only be developed through practice using the language. For example, according to Anderson's ACT theory (1983, 1993; Anderson & Lebiere, 1998), skill acquisition, including language acquisition, requires practice for the proceduralization of declarative knowledge and for the strengthening of learned productions. Turning to a model designed to apply specifically to language acquisition, in Bialystok's two-dimensional model of language acquisition (1994; Bialystok & Sharwood Smith, 1985), one dimension is the development of control of processing, that is, the ability to devote attentional resources efficiently. The development of second language fluency requires development along this dimension. However, other than looking at learners' response times in making their choices, this study is not really concerned with the development of second language fluency. Rather, it is concerned with the role of awareness in the acquisition of new second language features and in the conversion of input to intake. It was decided that the development of fluency was outside the scope of this paper and that it would have been impractical to study it with the number of learners involved.

A third limitation involves prior learning and the extent to which it was controlled. Extensive prior learning was indeed controlled by removing and replacing learners who had scored too high on the pretest. However, the pretest scores in Experiment One indicate that prior learning was still probably exerting an influence. Still, the low correlation between awareness levels and pretest scores in Experiment One indicates that this influence was not great. Things are somewhat more confused in Experiment Two. Based on the descriptive statistics of the pretest, prior learning seemed to be playing less of a role than in Experiment One. However, there was also a significant, if weak, correlation between pretest scores and awareness levels. When a study involves English, or other widely-studied natural languages, prior learning can only be controlled to a certain extent.

A final limitation of this study is that the pretest and posttest are a kind of grammaticality judgement test. Such tests may tap true linguistic competence, or they may only test learners' metalinguistic knowledge. Truscott (1998a) argues that studies of noticing do not provide evidence for the role of noticing in SLA since the tests used to measure acquisition may merely be tapping metalinguistic knowledge.

Implications

Keeping the limitations discussed above in mind, this study has implications for SLA research, SLA theory, and foreign and second language pedagogy. This final section discusses the implications of this study based on the results from Experiments One and Two discussed above.

Pedagogical implications. Most likely, the main benefit of explicit instruction is that it may facilitate noticing, and thus help in converting input to intake (Schmidt, 1995). It may also be that, for some learners, this help is not optional, it is required. However, since only some of the learners receiving explicit instruction in this study display a high level of awareness, this indicates that learners receiving equivalent amounts of explicit instruction in a classroom will nevertheless develop varying levels of awareness. Because of these differences, the facilitative effect of explicit classroom instruction will vary among learners. In addition, since learners receiving incidental instruction in this study also vary in their awareness, this indicates variability among learners in the effectiveness of classroom instruction intended to promote incidental learning. The main pedagogical implication of this study is that learners need to be given multiple opportunities, through a variety of instructional techniques, to develop awareness of features of the language. This implication is hardly earth-shattering, but it nevertheless needs to be remembered.

Implications for SLA research and theory. One implication for SLA research comes from the rather uninteresting fact that learners responded to items more quickly on the posttest than on the pretest. This result seems to be uninteresting from a theoretical point of view--the learners may just have wanted to complete the experiment. However, it is important from a methodological point of view. The pretest and posttest were rather long, containing 80 items each. Since the items were presented in a different random order to each learner, the desire to finish should not have caused response time differences between items within the same test. However, had the study employed tests in which items were not presented randomly, such as paper and pencil tests, there probably would have been shorter response times for later items. Researchers employing pretests and posttests in empirical studies of SLA need to keep in mind the potential influence of learner fatigue on the test results.

Another implication for SLA research is that researchers cannot provide learners with a certain kind of instruction and assume that it determines the learning processes that learners will use. For example, a hypothetical study which compared explicit instruction, input enhancement, and mere exposure may provide information about the effectiveness of these different approaches to instruction, and the explicit instruction and input enhancement may actually facilitate awareness. However, without an independent measure of awareness, such a study could not be claimed to be a study of the effects of awareness on learning. Similarly, demonstrations of learning without any explicit instruction should not be taken as evidence for learning without awareness. In addition, while it is not too difficult to develop a measure to assess different levels of awareness, researchers should recognize the methodological difficulty of establishing that learners are completely lacking in awareness. Unlike the measure used in this study, or in studies such as Robinson (1995a, 1996a, 1996b, 1997a) and Leow(1997), a measure capable of identifying zero-level awareness would need to meet completely the criteria of Shanks and St. John (1994) discussed in Chapter One.

Even though it is possible that the relationship found in this study between awareness and learning was not a causal relationship, by finding more learning among more-aware learners, the results support the argument that some level of awareness is necessary for learning, at least for those aspects of a second language which fall outside the scope of UG. The implication for SLA theory should be clear. If awareness does play a role in SLA, then theoretical models of SLA must take the role of awareness into account. Since this study, especially Experiment Two, finds a role for instance-based learning, this indicates that accumulation of instances should also be taken in to account in an SLA theory. Since the roles of awareness and instance accumulation in learning have mostly been investigated in the field of cognitive psychology, SLA theorists need to draw on theories and empirical evidence from cognitive psychology. This has, of course, already been done for some theoretical positions in SLA, such as the noticing hypothesis (Schmidt, 1990, 1993, 1994b, 1995; Robinson, 1995b), the fundamental similarity hypothesis (Robinson, 1995a, 1996a, 1996b, 1997a), and ideas about conceptually driven processing versus data-driven processing (Robinson, 1995b). However, other theoretical positions, in particular Krashen's (1981, 1982, 1994), seem to ignore findings from cognitive psychology and rely on the enigmatic workings of a LAD to explain how languages are acquired. Unfortunately, such theoretical positions do nothing to illuminate the cognitive processes involved in SLA.

[1]Although the researcher did not keep a record of his Japanese learning experience, it is his subjective opinion that he had an experience similar to this hypothetical learner. Japanese is a language which does not require subjects to be stated. In fact, not omitting subjects often results in unnatural-sounding language production. The researcher learned Japanese in a mixed instructional and naturalistic setting, but does not remember ever being explicitly informed that he could omit subjects. However, after being in Japan a few months and attempting to interact in Japanese on an almost daily basis, the researcher noticed that, while he always used *watashi* or *boku* (both mean *I*) when stating what he wanted to do, what he liked, what he did, etc. . . , Japanese speakers almost never used these words. The researcher then began consciously to drop *watashi* and *boku* in first person singular contexts. Although the researcher had the concept of grammatical subject, it was only much later that he realized that all subjects, not just the first person singular subject pronoun, could usually be dropped in Japanese. In other words, he initially noticed that *watashi* and *boku* could be dropped without having to refer to the concept of grammatical subject.

[2]In ACT*, separate procedural rules, or productions, could be made more efficient through a process of composition, the joining of separate productions into a single new production. As explicitly pointed out in Anderson (1993), the process of composition has been dropped from ACT-R. Instead, multiple productions can deposit a new example in declarative memory, which may then be used by analogy in the compilation of a single new production.

[3]There is, I believe, an ethical problem with the Japanese study. The participants in the study are learners of Japanese who, presumably, desire to learn the language. They are not participants in a study using an artificial language for whom it is unimportant whether they actually learn to communicate in the language. Even so, learners are taught a structure, the correct ordering of adjectives, which does not actually exist in Japanese. For example, in Japanese, both *akakute ookii hako* (red large box) and *ookikute akai hako* (large red box) are grammatically correct. It seems to be a disservice to the learners to teach them something which is not only useless but also untrue.

[4]All learners who participated in this study were students at an English language school at the time of their participation. It is the researcher's opinion that they form a single population.

Appendix A—Place Names Used on Tests and in Instruction, Experiment One

Place Names Used in Instruction and on Both Tests

Schools:
Kaimuki Elementary School
George Washington High School

Countries:
Canada
France

Universities:
West Virginia University
Arizona University

Parks:
Kapiolani Park
Central Park

Single Mountains:
Mount Fuji
Mount Everest

Parts of Continents:
Central Africa
Western Australia

Hospitals:
General Hospital
Memorial Hospital

Single Lakes:
Lake Michigan
Lake Erie

Bays:
Hanauma Bay
Kaneohe Bay

Single Islands:
Easter Island
Saipan

Countries with *the*:

the United Kingdom
the Republic of China

Universities with *the*:
the University of Hawai‘i
the University of Florida

Tourist Attractions:
the Washington Monument
the Eiffel Tower

Regions:
the East Coast
the Middle East

Island Chains:
the Hawaiian Islands
the Falkland Islands

Hotels:
the Hilton Hotel
the Royal Hawaiian Hotel

Mountain Chains:
the Alps
the Himalayas

Rivers:
the Yellow River
the Nile River

Peninsulas:
the Iberian Peninsula
the Florida Peninsula

Associations:
the Young Women's Christian Association
the Hawaiian Association of Language Teachers

Place Names Used Only on Pretest

Schools:
Roosevelt Middle School
Oakwood Elementary School

Countries:
Russia
Italy

Universities:
Hawaiʻi Pacific University
Boston University

Parks:
Ala Wai Park
Ueno Park

Single Mountains:
Mount McKinley
Mount Kilimanjaro

Parts of Continents:
Southeast Asia
Central America

Hospitals:
Duke Hospital
Central Hospital

Single Lakes:
Lake Geneva
Lake Ontario

Bays:
San Francisco Bay
Tokyo Bay

Single Islands:
Tahiti
Molokai

Countries with *the*:
the Republic of Ireland
the United States

Universities with *the*:
the University of Iowa
the University of South Carolina

Tourist Attractions:

the Golden Gate Bridge
the Sphynx

Regions:
the Northwest
the Southeast

Island Chains:
the Marshall Islands
the Moldive Islands

Hotels:
the Outrigger Hotel
the Kahala Mandarin Hotel

Mountain Chains:
the Rocky Mountains
the Andes

Rivers:
the Mississippi River
the Jordan River

Peninsulas:
the Izu Peninsula
the Yucatan Peninsula

Associations:
the Young Buddhist Association
the American Association of Retired People

Place Names Used Only on Posttest

Schools:
John Adams Junior High School
McKinley High School

Countries:
Mexico
Egypt

Universities:
Oxford University
Harvard University

Parks:
Maple Park
Old Stadium Park

Single Mountains:
Mount Etna
Mount Saint Helens

Parts of Continents:
South Asia
North Africa

Hospitals:
Shriner's Hospital
Presbyterian Hospital

Single Lakes:
Lake Norman
Lake Victoria

Bays:
Boston Bay
Charleston Bay

Single Islands:
Maui
Kauai

Countries with *the*:
the Republic of Korea
the United States of Mexico

Universities with *the*:
the University of California
the University of Kentucky

Tourist Attractions:
the Brooklyn Bridge
the Pyramids

Regions:
the West Coast
the South

Island Chains:

the Canary Islands
the Shetland Islands

Hotels:
the Ala Moana Hotel
the Sheraton Hotel

Mountain Chains:
the Ural Mountains
the Appalachian Mountains

Rivers:
the Thames River
the Amazon River

Peninsulas:
the Arabian Peninsula
the Korean Peninsula

Associations:
the National Association for the Advancement of Colored People
the National Basketball Association

Appendix B—Rules of Thumb for Article Usage, Experiment One

Do not use "the" with the names of schools.

Do not use "the" with the names of most countries.

Do not use "the" with the names of universities that do not include the word "of."

Do not use "the" with the names of parks.

Do not use "the" with the names of single mountains.

Do not use "the" with the names of parts of continents.

Do not use "the" with the names of hospitals.

Do not use "the" with the names of single lakes.

Do not use "the" with the names of bays.

Do not use "the" with the names of single islands.

Use "the" with the names of tourist attractions.

Use "the" with the names of regions not including a continent or country name.

Use "the" with the names of island chains.

Use "the" with the names of hotels.

Use "the" with the names of mountain chains.

Use "the" with the names of countries that include words such as "republic," "united," or "of."

Use "the" with the names of universities that include the word "of."

Use "the" with the names of rivers.

Use "the" with the names of peninsulas.

Use "the" with the names of associations.

Appendix C--Questions Used on Delayed Posttest

1. What is the highest mountain in the world?

2. Where is Mount Everest? Do you know the name of the mountain range?

3. What is the highest mountain in Japan?

4. What is the longest river in the world?

5. Do you know the longest river in America?

6. Can you name the countries in North America?

7. What is the name of this island?

8. What are some of the other islands around here?

9. These islands together make a group of islands. What's the group of islands' name?

10. Do you know the names of any universities in Hawai‘i?

11. Can you name any hotels in Waikiki?

12. Can you name any good places to go snorkeling on Oahu?

13. What part of the United States in California in?

14. What part of the United States in New York in?

15. Can you name any lakes in the world?

16. Can you name any peninsulas in the world?

17. Do you know the names of any schools in Honolulu, like elementary schools, or junior high schools, or high schools?

18. Where is Thailand?

Appendix D--Sample Transcript of Delayed Posttest

R=Researcher
L=Learner
Note: Tokens where L just repeats after R are not counted. Self repetitions of a token by L are not counted either.

R: If you don't know the answer to a question, just say I don't know.
L: Okay
R: What is the highest mountain in the world?
L: . . . Mount Everest.
R: Where is Mount Everest?
L: Tibet.
R: Okay. Well, Nepal. Do you know the name of the, uh, mountain range, that it's part of?
L: No.
R: Oh, the Himalayas.
L: Mm.
R: What's the highest mountain in Japan?
L: Mount Fuji.
R: Yeah. What's the longest river in the world?
L: . . . I don't know.
R: Oh, probably the Nile. Well, the Nile or the Amazon. Do you know the longest river in America?
L: I don't know.
R: Oh, the Mississippi.
L: Mm.
R: Uh, do you, can you name the countries in North America?
L: . . . North America? Countries? (yeah) uh, like New Jersey?
R: Those, those are states. (oh) The whole country.
L: The United States.
R: Okay. Any others?
L: Alaska.
R: Well, that's part of the United States. Canada (Canada) or Mexico. Uh . . .
L: Mexico is North America?
R: Yeah. We consider that in North America.
L: Oh, really?
R: Yeah. Uh, what is the name of this island?
L: Hawai'i Island.
R: This island?
L: Oh! Oahu Island.
R: Okay. (ha-ha) What are some other islands around here?
L: . . . Lanai Island, Big Island, Maui Island . . .
R: Okay. These islands together make a group of islands. What's the group of islands' name?

L: Hawai'i.
R: Okay. Also the Hawaiian Islands
L: The Hawaiian . . .
R: Uh, do you know the names of any universities in Hawai'i?
L: UH.
R: What's the full name?
L: University of Hawai'i.
R: Okay, do you know any others?
L: KCC . . . uh, Kapiolani Communications College.
R: Community College.
L: Community College.
R: ICC is Intercultural Communications College. Those are community colleges. Okay, uh, can you name any hotels in Waikiki?
L: The Hilton Hawaiian Village, the Hyatt Regency, . . . the Hyatt Regency . . . many many.
R: Okay. A very different question, the next question, uh, can you name any good places to go snorkeling in, on Oahu?
L: Mm, Hanauma Bay.
R: Okay. And a very different question. What part of the United States is California in?
L: . . . West.
R: Yeah, the West Coast.
L: The West Coast.
R: What part of the United States is New York in?
L: The East Coast.
R: Okay. Can you name any lakes in the world?
L: Lakes? . . . Victoria Lake?
R: Okay. (I don't know.) Can you name any peninsulas in the world?
L: . . . Korean Peninsula.
R: Okay. And, do you know the names of any schools in Honolulu, like elementary schools, or junior high schools, or high schools?
L: Kaimuki High School.
R: Okay. And one more question. Where is Thailand?
L: I know, in Asia.
R: What part of Asia?
L: I don't know.
R: Okay, Southeast Asia.
L: Southeast.
R: Very good.

Appendix E--Replacement Items Used in Experiment Two

Pretest, Instruction, and Posttest Items:

Libraries:
McCully Library
Hamilton Library

Deserts:
the Sahara Desert
the Painted Desert

Pretest:

Libraries:
Sinclair Library
Kaneohe Library

Deserts:
the Mojave Desert
the Gobi Desert

Posttest:

Libraries:
Metropolitan Library
Manoa Library

Deserts:
the Arabian Desert
the Atacama Desert

References

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259-302). Honolulu, HI: University of Hawai'i Press.

Allwright, D. (1984). Why don't learners learn what teachers teach? The interaction hypothesis. *Language learning in formal and informal contexts: Proceedings of a joint seminar of the Irish and British Associations for Applied Linguistics,* 3-18.

Anderson, J. R. (1983). *The architecture of cognition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). NY: Freeman.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 932-945.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16,* 157-168.

Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second-language acquisition. *Applied Linguistics, 6,* 101-117.

Brown, J. D. (1988). *Understanding research in second language learning.* Cambridge: Cambridge University Press.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal, 79,* 179-193.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15,* 357-386.

Celce-Murcia, M. & Larsen-Freeman, D. (1983). *The grammar book: An ESL/EFL* *teacher's course.* Boston: Heinle & Heinle Publishers.

Clark, R. E. & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science, 280,* 77-82.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers, 23,* 257-271.

Curran, T. & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 189-202.

de Graaff, R. (1997). The Experanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19,* 249-276.

DeKeyser, R. M. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *The Modern Language Journal, 77,* 501-513.

DeKeyser, R. M. (1994). How implicit can adult second language learning be? *AILA Review, 11,* 83-96.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17,* 379-410.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*, 195-221.

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of ESL relativization. *Studies in Second Language Acquisition, 13,* 431-469.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197-261). Cambridge, UK: Cambridge University Press.

Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology, 5,* 289-318.

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16,* 409-441.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning, 44,* 449-491.

Fotos, S. S. (1993). Consciousness raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics, 14,* 385-407.

Fotos, S. S. (1994). Integrating grammar instruction and communicative language use through grammar conscioiusness-raising tasks. *TESOL Quarterly, 28,* 323-351.

Green, P. S. & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics, 13,* 168-184.

Guilford, J. P., & Fruchter, B. (1973). *Fundamental statistics in psychology and education* (5th ed.). New York, NY: McGraw-Hill.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* Boston, MA: Heinle & Heinle Publishers.

Hewson, J. (1972). *Article and noun in English.* The Hague: Mouton.

Hulstijn, J. L. (1996). Not all grammar rules are equal: Giving grammar instruction its proper place in foreign language teaching. In R. Schmidt (Ed.), *Attention & awareness in foreign language learning* (pp.359-386). Honolulu, HI: University of Hawai'i Press.

Hulstijn, J. H. & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*, 97-112.

Hulstijn, J. H., Hollander, M., Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurance of unknown words. *The Modern Language Journal, 80,* 327-339.

Izumi, S., & Lakshmanan, U. (1998). Learnability, negative evidence and the L2 acquisition of the English passive. *Second Language Research, 14,* 62-101.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19,* 357-377.

Jourdenais, R. (1996). The limitations of think-alouds. Paper presented at the American Association of Applied Linguistics, Chicago, IL.

Kachru, B. B., & Nelson, C. L. (1996). World Englishes. In S. L. McKay & N. H.

Hornberger (Eds.), *Sociolinguistics and language teaching* (pp. 71-102). Cambridge: Cambridge University Press.

Kaluza, H. (1981). *The use of articles in contemporary English.* Heidelberg: Julius Groos Verlag.

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal, 75,* 305-313.

Klein, W. (1986). *Second language acquisition.* Cambridge: Cambridge University Press.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Oxford: Pergamon Press.

Krashen, S. D. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp.45-78). Great Britain: Academic Press Limited.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47,* 467-505.

Leow, R. P. (1998). The effects of amount and type of exposure on adult learners' L2 development in SLA. *The Modern Language Journal, 82*, 49-68.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22,* 1-35.

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology, 105,* 317-339.

Logan, G. D. & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1022-1050.

Long, M. L., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in

SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal, 82,* 357-371.

Master, P. (1990). Teaching the English articles as a binary system. *TESOL Quarterly, 24,* 461-478.

Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229-252). UK: Cambridge University Press.

McLaughlin, B. (1990). "Conscious" versus "unconscious" learning. *TESOL Quarterly, 24,* 617-634.

Mondria, J., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics, 12,* 249-267.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324-354.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419). Great Britain: Academic Press Limited.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology, 118,* 219-235.

Riggenbach, H. & Samuda, V. (1997) *Grammar dimensions 2: Form, meaning, and use.* Second Edition. D. Larsen-Freeman (Series Director). Boston: Heinle & Heinle Publishers.

Robinson, P. (1995a). Aptitude, awareness, and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.), *Attention & awareness in foreign language learning* (pp. 303-358). Honolulu, HI: University of Hawai'i Press.

Robinson, P. (1995b). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45,* 283-331.

Robinson, P. (1996a). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition, 18,* 27-68.

Robinson, P. (1996b). *Consciousness, rules, and instructed second language acquisition.* New York: Peter Lang.

Robinson, P. (1997a). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47,* 45-99.

Robinson, P. (1997b). Generalizability and automaticity of second language learning under implicit, incidental, enhanced and instructed conditions. *Studies in Second Language Acquisition, 19,* 223-247.

Robinson, P., & Ha, M. A. (1993). Instance theory and second language rule learning under explicit conditions. *Studies in Second Language Acquisition, 15,* 413-438.

Rose Chang, K., & Flint Smith, W. M. (1991). Cooperative learning and CALL/IVD in beginning Spanish: An experiment. *The Modern Language Journal, 75,* 205-211.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629-639.

Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 42-55.

Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition*. Rowley, MA: Newbury House, pp. 237-274.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129-158.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14,* 357-385.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13,* 206-226.

Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11,* 11-26.

Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165-209). London: Academic Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention & awareness in foreign language learning* (pp. 1-64). Honolulu, HI: University of Hawai'i Press.

Schmidt, R. & Frota, S. N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Schwartz, B. (1986). The epistemological status of second language acquisition. *Second Language Research, 2,* 120-159.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15,* 147-164.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367-447.

Slimani, A. (1992). Evaluation of classroom interaction. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 197-220). Cambridge: Cambridge University Press.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13,* 138-163.

Tomlin, R. & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16,* 183-203.

Trahey, M. (1996). Positive evidence in second language acquisition: Some long term effects. *Second Language Research, 12,* 111-139.

Truscott, J. (1998a). Noticing in second language acquisition: A critical review. *Second Language Research, 14,* 103-135.

Truscott, J. (1998b). Instance theory and Universal Grammar in second language research. *Second Language Research, 14,* 257-291.

VanPatten, B. & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15,* 225-243.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research, 7,* 133-161.

White, L., Spada, N., Lightbown, P. M., Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics, 12,* 416-432,

Whittlesea, B. W. A., & Wright, R. L. (1997). Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 181-200.

Winitz, H. (1996). Grammaticality judgment as a function of explicit and implicit instruction in Spanish. *The Modern Language Journal, 80,* 32-46.

Winter, B. & Reber, A. (1994). Implicit learning and the acquisition of natural languages. In N. Ellis (Ed.), *Implicit and explicit learning of languages,* (pp. 115-145). London: Academic Press.

Yang, L. R., & Givon, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition: The Keck second language learning project. *Studies in Second Language Acquisition, 19,* 173-193.

Zimmerman, C. B. (1997). Do reading and interactive vocabulary instruction make a difference? An empirical study. *TESOL Quarterly, 31,* 121-140.

Zobl, H. (1995). Converging evidence for the 'acquisition-learning' distinction. *Applied Linguistics, 16,* 35-56.

ABSTRACT

Explicit instruction can facilitate learner awareness of the surface features of a language, but does not guarantee it. Similarly, learners in an incidental learning condition are not necessarily unaware. This study investigated the development of awareness, among Japanese ESL learners, of rules of thumb for the use of zero and definite articles with place names under an explicit instruction condition, in which learners were given the rules plus examples, and an incidental instruction condition, in which learners responded to sentences containing examples. All instruction was computerized. Instruction was given in English and was followed by a twenty-question debriefing interview conducted in the learners' L1 in order to assess their awareness. The findings show that awareness could develop under either condition, but that the explicit condition was much more facilitative. The study also found a very strong relationship between awareness and improved learner performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One
The Role of Awareness in Learning and Acquisition

Introduction

The role of consciousness in second language acquisition (SLA) continues to divide researchers, with some claiming that it plays no role in the acquisition of linguistic competence and others claiming that there can be no acquisition without it. Arguments about consciousness are complicated by the imprecision of the terms *conscious* and *unconscious* , limiting the usefulness of these terms. In addition, while researchers who see no role for consciousness in language acquisition generally argue that the acquisition of language is qualitatively different from other kinds of learning, researchers in the field of cognitive psychology, who have also been debating the role of consciousness in learning, generally do not give language acquisition such unique status.

This chapter discusses the different theoretical positions on one aspect of consciousness, awareness, during SLA, examines theories of learning in cognitive psychology, and looks at the relationship between cognitive psychology and SLA. Chapter Two discusses how effects of instruction studies have and have not examined the role of awareness. Chapters Three and Four describe two empirical experiments investigating the role of awareness. Finally, Chapter Five discusses the results and implications of the experiments.

Definitions

Both McLaughlin (1990) and Schmidt (1994a) have described the imprecision of the terms *conscious* and *unconscious* and have proposed more precise terminology to describe the different meanings that these terms can refer to. There can, though, also be confusion over ostensibly more precise terms, such as *explicit* and *implicit*. In an effort to avoid confusion, the terms *conscious* and *unconscious* will mostly be avoided in this paper and the following definitions will be used for other terminology.

First, based on the recommendations of Schmidt (1994a), the term *incidental* is used to refer to the learning of features which learners do not intend to learn (Schmidt’s recommendation 1). In the third chapter, the term *incidental instruction,* rather than *incidental learning*, will be used since it is believed that some learners receiving incidental instruction may actually be learning intentionally. *Explicit* is used to refer to learning and instruction which involve the provision of rules (Schmidt’s recommendation 6). Again, in the third chapter, it is possible that learners will not learn anything from the rules and examples provided to learners, so the term *explicit instruction* is used rather than *explicit learning.*

Next, *awareness* is defined, or rather operationalized, as the ability of learners to verbalize, in their first language (L1), knowledge of instructional targets soon after the instruction has been completed. This knowledge may be expressed as either rules or examples. It is quite possible that there are aware learners who are unable to verbalize this awareness (Schmidt, 1994a), and McLaughlin (1990) has described "learning with awareness" (p.628) and "ability to report what is known" (p. 628) as separate aspects of the term *conscious*. However, for the purposes of this paper, learners who are unable to verbalize are described as *less- aware* while learners who are able to verbalize are described as *more-aware*. It does not seem possible, at the present time, to assess awareness directly and, therefore, the researcher will rely on the indirect measure of ability to verbalize. The instrument used to assess awareness in this study is described in the third chapter.

Implicit learning (Reber, 1989; Winter & Reber, 1994) is defined as the ability to induce knowledge of rules underlying a complex stimulus domain by focussing attention on the surface features of that domain. *Tacit knowledge* refers to the knowledge base resulting from such learning. Neither implicit learning nor tacit knowledge necessarily entail a lack of awareness of surface features during learning. What learners are unaware of is the rule system that they are believed to have induced from regularities in the surface features as well as the process of inducing this system.

Finally, *subliminal learning* is used to describe learning which is supposed to occur in the complete absence of any awareness of what is being learned. It should be noted here that, along with Schmidt (1990, 1993, 1994b, 1995), the author is skeptical of any role for subliminal learning in language acquisition.

Language Acquisition and Awareness

Acquisition without awareness. Arguments that SLA occurs without awareness generally draw a distinction between learning and acquisition. For example, according to Krashen (1981, 1982, 1994), language learning is a conscious process in which learners learn and are aware of facts about the language, while acquisition is incidental and implicit in that learners acquire language while they are focused on communication and they do not become aware of facts about the language. Krashen, and often others who draw this distinction between learning and acquisition, also takes a strong noninterface position—he claims that learning can never be transformed into acquisition. Therefore, if Krashen is correct, explicit instruction should be of almost no benefit to people attempting to acquire a second language. Also, taking an explicit approach to learning a language, that is, trying to figure out rules, trying to apply conscious knowledge, focusing attention on form, should be of almost no benefit as well. The only benefits explicit instruction or explicit learning strategies can provide are 1) giving learners a rather clumsy set of tools for monitoring their output when conditions allow and 2) providing learners with samples of the language which may possibly contain elements which can be acquired.

Paradis (1994) also argues in support of learning without awareness. Paradis, echoing Krashen, claims that only "implicit linguistic competence" can be used automatically, that grammatical knowledge can be used to check correctness but not as "part of the automatic production process," and that this "metalinguistic knowledge" cannot be proceduralized, or converted into linguistic competence. For Paradis, developing linguistic competence means developing automaticity in the use of the language, and "what becomes automatic is not what the learners focus their attention on, or are even aware of" (p. 402). Paradis supports his claims with evidence from neurolinguistics, which supposedly shows that different parts of the brain are responsible for the use of implicit linguistic competence and conscious grammatical knowledge of which learners are aware.

Some researchers who study SLA within a Universal Grammar (UG) framework also support the distinction between learning and acquisition and the claim that the acquisition of linguistic competence occurs in the absence of awareness. Zobl (1995) argues that UG provides theoretical support for the learning/acquisition distinction and claims that empirical evidence shows that acquisition but not learning is uniform across learners, that acquisition is resistant to forgetting, that explicit metalinguistic knowledge is dissociated from learner performance, and that natural languages are too complex to be acquired by general, non-linguistic learning mechanisms. Schwartz (1986, 1993) draws a distinction, similar to Krashen's, between linguistic competence, which is based on UG and acquired through a language specific language acquisition device (LAD), and learned linguistic knowledge (LLK), which is learned explicitly using general learning mechanisms. Linguistic competence can only be developed on the basis of positive evidence in the input. Negative evidence about what is not possible, explicit instruction, and error correction can result in LLK, but have no impact on linguistic competence. The acquisition of LLK may require awareness but the acquisition of linguistic competence does not. However, her position is somewhat different from the positions of Krashen and Paradis in that LLK may play an important role in helping second language speakers, if they have a highly developed, but ultimately unnative-like, linguistic competence and plenty of practice, display native-like behavior. The use of LLK does not have to be slow and clumsy.

Finally, Bialystok (1994; Bialystok & Sharwood Smith, 1985) draws a completely different distinction between two orthogonal dimensions of language acquisition, one of *analysis of knowledge* and one of *control of processing*. Unlike the learning/acquisition distinction, development along both these dimensions is necessary for the acquisition of linguistic competence. Awareness is not seen as necessary for development along either dimension.

It is not completely clear in Krashen's writings whether what he calls the subconscious process of language acquisition should be called implicit or subliminal. Krashen (1994) calls the process “implicit” and “incidental,” but implicit, as defined above, means the unconscious induction of rules based on what is attended to, and incidental means language learning without the intention to learn while focusing on meaning. Neither implicit nor incidental language learning necessarily take place in the absence of awareness of surface features, and for language learning to be implicit, it need only occur in the absence of awareness of the rules underlying noticed surface features and the induction of such rules. However, with his input hypothesis, affective filter hypothesis, and acquisition-learning hypothesis, Krashen strongly implies that acquisition is unconscious in the sense of occurring in the total absence of awareness. According to the input hypothesis (Krashen, 1994), input becomes available to the LAD, that is, becomes intake, by being comprehensible and by being at the *i+1* level, that is, by being just above the learner’s current level. Assuming that the learner’s affective filter is low, *i+1* input will then be used for acquisition. In this scheme, awareness, attention, etc., play absolutely no role. According to the acquisition-learning hypothesis, learned knowledge, which seems to mean both that the knowledge was explicitly learned and is held in explicit memory, can never become acquisition, so in fact it is impossible for awareness to play any role. It seems that what Krashen, as well as Paradis and other supporters of acquisition without awareness, is arguing for is not implicit language acquisition so much as subliminal language acquisition, that is, acquisition without any attention to or awareness of any aspect of what is acquired.

No acquisition without awareness. On the other hand, according to Schmidt's noticing hypothesis (1990, 1993, 1994b, 1995; called the noticing the gap principle in Schmidt & Frota, 1986), awareness at the level of noticing is necessary to convert input into intake which can be used for SLA. Robinson (1995b) also supports a role for noticing in SLA and precisely defines "noticing as detection with awareness and rehearsal in short-term memory" (p. 318). The noticing hypothesis does not imply that learners call on explicitly held rules during fluent second language use or that learners necessarily have episodic or recognition memory of what they have learned. Rather, it means that a certain level of awareness of language features exists at the time that those features are acquired. Also, while the noticing hypothesis states that awareness is necessary and sufficient for converting input to intake, it does not imply that awareness is sufficient for SLA. As is obvious from the number of people who have learned and are aware of facts about a language yet cannot communicate in it, input and interaction are necessary for the development of fluency in a second language (Schmidt, 1995). According to the noticing hypothesis, unconscious language learning, that is, SLA without awareness, or subliminal SLA, is not possible.

Schmidt (1995) also points out a basic flaw in the argument that the complexity of languages requires that they be learned unconsciously (e.g., Krashen, 1994; Paradis, 1994), since conscious acquisition of something as complex as a natural language would be impossible. While, for example, the general failure of instructed ESL learners to master the English article system indicates the insufficiency of instruction, "the typical failure of both instructed and uninstructed learners in these areas of grammar counts even more heavily against arguments for the success of implicit learning" (p. 40). The general lack of people who have successfully and completely acquired a second language as adults may be the result of the complexity of natural languages, but this does not mean that unconscious learning provides a way around this complexity. As DeKeyser (1994) points out, it is very difficult, maybe even impossible, to find people with a high level of second language proficiency who have acquired this second language as adults and who have never received any explicit instruction nor ever adopted any explicit learning strategies.

Unfortunately, the noticing hypothesis can probably be neither proven nor falsified (Schmidt, 1995). As Schmidt states, awareness at the time of learning "is fleeting and cannot be completely recorded" (p. 28). Therefore, it cannot be shown that learning always occurs with awareness or that learning can occur in the complete absence of awareness. However, it should be possible, by gathering in-depth evidence of learner awareness shortly after learning, to compare the learning of more-aware and less-aware learners. If the learning of more-aware learners is superior to that of less-aware learners, this would support, though of course not prove, the noticing hypothesis. If the learning of more-aware learners is not superior, or even inferior, this would provide counter-evidence to the noticing hypothesis, even though it would not disprove it. (Of course, in either scenario, the measure of awareness would have to meet certain criteria. See below.)

Robinson (1995a, 1996a, 1996b, 1997a) has proposed the fundamental similarity hypothesis, which claims that all adult second language learning is based on the same conscious mechanisms. What may appear to be unconscious implicit or incidental learning still involves attention to, noticing of, and rehearsal in memory of the linguistic features which are learned. In other words, implicit and incidental learning rely on conscious processes. While the conscious mechanisms underlying SLA are fundamentally similar, different types of instruction and task demands can encourage different varieties of conscious processing. As Robinson (1996b) states:

> . . . all L2 learning is the result of conscious awareness of form at the point of input, which is explicitly attended . . . (and) the effects of learning under implicit and explicit conditions are artifacts of the variety of conscious processing encouraged by different task demands. (p. 180)

Obviously, the claim that L2 learning results from “awareness of form at the point of input” is identical to the noticing hypothesis of Schmidt discussed above. As will be discussed in chapter two, Robinson has found empirical support for the fundamental similarity hypothesis in general and for the causal role of awareness in SLA in particular.

Finally, DeKeyser (1994) argues that there is a tendency in SLA research “to equate implicit vs. explicit knowledge with knowledge acquired implicitly vs. explicitly” (p. 85). This is similar to Schmidt (1995), who points out that the existence of implicit knowledge is irrelevant to the question of awareness at the point of learning. DeKeyser (1994) argues as well that studies investigating the knowledge used by advanced L2 speakers “do not shed any direct light . . . on the relationship between the two types of knowledge at the time a specific rule is learned” (p. 85). In other words, lack of conscious awareness of L2 rules on the part of fluent and accurate speakers does not entail that the learners were not consciously aware of rules or of formal features of the L2 when they first learned them.

Tomlin and Villa (1994) argue that it is detection, a component of attention, and not awareness which is necessary for acquisition. In their view, detection can be facilitated by the other two components of attention—alertness and orientation—and these other components can be facilitated by awareness. Awareness, therefore, can indirectly facilitate detection and acquisition, but detection is possible without awareness, so awareness is not necessary to convert input into intake. Unlike Krashen (1981, 1982, 1994), Paradis (1994), and others who argue for two learning systems, though, Tomlin and Villa do not argue for two learning systems, one requiring awareness and the other not. Instead, they argue that detection and awareness can be dissociated and that it is detection, not awareness, which is necessary for learning to take place. However, Tomlin and Villa admit, in an endnote, that there are problems in separating detection and awareness empirically:

> Cognitive registration of information has been dissociated from awareness, but detection of targets has most often been assessed via a response that implicates awareness. (p. 199)

Schmidt (1995) points out that the studies cited by Tomlin and Villa as evidence for a dissociation between detection and awareness actually show "a *very* strong relationship between awareness and learning" (p. 21) and "more learning with more noticing and less learning with less noticing" (p. 22). While Schmidt admits that detection and awareness can be dissociated, and that there is evidence that certain kinds of perception, that is, subliminal perception and blindsight perception, are possible without awareness, he argues that learning without awareness has not been demonstrated, that awareness is good for learning, and "that focal attention and awareness are essentially isomorphic" (p. 20).

Truscott (1998a) criticizes the noticing hypothesis. First, he claims, correctly, that a weak version of the noticing hypothesis, that simple awareness of the existence of input is necessary for SLA, is uninteresting. He then claims that the strong version, that awareness of grammatical details is necessary for SLA, is not supported by empirical evidence from cognitive psychology and is not clear about what exactly needs to be noticed. Truscott claims that implicit learning studies in the field of cognitive psychology allow one "safely (to) conclude that the evidence does not show that awareness of the information to be acquired is necessary for learning" (p. 109). Given that Truscott admits that "no firm conclusions can be drawn about implicit learning" (p.109), it is somewhat surprising that he is able safely to reach this conclusion. Specifically, it is not clear whether sequence learning demonstrated under implicit conditions results from a tacit knowledge of rules gained in the absence of awareness or from memory of allowable sequences (Shanks & St. John, 1994). Also, since implicit learning as operationalized in cognitive psychology does involve attention to and awareness of surface features, it is not incompatible with the noticing hypothesis.

Truscott's other criticism, that it is not clear exactly what details need to be noticed, is valid—empirical investigations of the noticing hypothesis need to specify what learners are becoming aware of. However, Truscott's argument that, for example, noticing the absence of subject pronouns in Spanish requires the concept of grammatical subject is incorrect. Conceivably, a native speaker of English, with no explicit concept of subject, learning Spanish naturalistically in a Spanish-speaking country, could notice that Spanish-speakers often seem to drop *yo* when talking about things they do themselves. This hypothetical learner would not need the concept of grammatical subject in order to notice this and would not need to be aware of a rule such as, *You can drop subject pronouns*.[1]

Truscott argues for a reformulation of the noticing hypothesis, claiming that noticing is necessary for the acquisition of metalinguistic knowledge but not for SLA. Such a reformulation would make the noticing hypothesis compatible with Krashen's learning/acquisition distinction in that noticing would be necessary for learning but not for acquisition. In making this reformulation, Truscott points out that empirical studies of the noticing hypothesis have generally assessed learning through tests which may tap metalinguistic knowledge rather that linguistic competence. This is also a valid criticism—investigations of the role of awareness in SLA need to use assessment measures that are less likely to rely solely on metalinguistic knowledge. However, if such measures are developed, then the noticing hypothesis can be investigated without such a reformulation.

Learning and Awareness in Cognitive Psychology

In general, researchers in SLA, particularly those who support language acquisition without awareness, believe that language acquisition is fundamentally different from other kinds of learning. For example, Schwartz (1986) argues that there is overwhelming evidence that first language acquisition is different from other kinds of learning and that the null hypothesis in SLA research should be that first and second language acquisition rely on the same mechanisms. This null hypothesis thus entails that second language acquisition is also qualitatively different from other kinds of learning. In the field of cognitive psychology, however, common mechanisms are generally assumed to underlie all higher cognitive functions, including language acquisition (e.g., Anderson, 1983, 1993; Reber, 1989). Rather than looking at possible differences in the learning of knowledge from different domains, research and debates in the field of cognitive psychology tend to concern distinctions between explicit and implicit learning, between attentional and non-attentional learning, and between instance-based and rule-based learning. This section discusses some evidence for and against these distinctions and of how they relate to SLA. This section does not, though, provide a comprehensive review of the field of cognitive psychology, as this would be beyond the scope of this paper.

Explicit and implicit learning. Reber is one of the most prolific researchers in cognitive psychology claiming a dissociation between learning and awareness. In a review of his own and other's work on implicit learning and the resulting tacit knowledge base (Reber, 1989), Reber claims that learning the underlying rules of a complex stimulus domain occurs implicitly, in the absence of awareness of the rules, and results in abstract, structured knowledge, a tacitly held system of rules, which also exists in the absence of awareness. However, this does not mean that implicit learning takes place without attention being directed to the stimulus domain—implicit learning is "a by-product of the application of attention to relevant rule-governed structures in the environment" (Winter & Reber, 1994, p. 117). In other words, attention is given to surface features, implicating awareness of those features at the point of learning.

Reber's research paradigm is not directly related to SLA, but superficially his position would seem to support the positions of Krashen and Paradis, especially since Reber (1989; Winter & Reber, 1994) claims that implicit learning has a distinct advantage over explicit learning—implicit learning can, and will, take place when the rules underlying a complex stimulus domain cannot be found, while explicit learning can only happen when such rules can be found. This is similar to Krashen's assertion (1981, 1982, 1994; also Zobl, 1995) that only simple second language rules can be consciously learned while more complex rules can only be acquired. However, while both Krashen and Paradis take a strong non-interface position, arguing that learning can never be converted into acquisition, in Reber's view, it is the attending to and memorization of surface features of a complex domain (that is, the memorization of sequences) that leads to unconscious induction of the rules underlying the domain. In fact, implicit learning in the experimental work of Reber is operationalized by instructing learners to memorize sequences generated by a complex underlying set of rules. Thus learners are aware of surface features but are unaware of rules either at the time of learning or when they are used. According to Reber (1989), for the "sophisticated processes (which) are more uniquely the stuff of humanity" (p. 232), some tacit knowledge can be "derived from explicit processes that (become) automatic only after pained, conscious action" (p. 232). Another difference with Krashen is that Krashen (1981, 1982, 1994), by giving a major role in SLA to a LAD, takes a nativist position in which there exists a mental organ specifically for acquiring language, a position also taken by Schwartz (1986, 1993) and Zobl (1995). Reber (1989; Winter & Reber, 1994), on the other hand, argues that the process of implicit learning is a generalized cognitive process of interacting with the environment and is not unique to any one domain.

Not everyone accepts that Reber's work provides supporting evidence of the existence of implicit learning. In a review of the evidence for and against the existence of unconscious learning, Shanks and St. John (1994) conclude that the existence of learning without awareness has not been adequately demonstrated. One of their main criticisms of empirical studies which are claimed as evidence for learning without awareness is the instruments used to measure participants' awareness. They give two criteria, the information criterion and the sensitivity criterion, for deciding whether a particular measure of awareness is adequate:

> *Information Criterion:* Before concluding that subjects are unaware of the information that they have learned and that is influencing their behavior, it must be possible to establish that the information the experimenter is looking for in the awareness test is indeed the information responsible for performance changes. (p. 373)
>
> *Sensitivity Criterion:* To show that two dependent variables (in this case, tests of conscious knowledge and task performance) relate to dissociable underlying systems, we must be able to show that our test of awareness is sensitive to all of the relevant conscious knowledge. (p. 373)

Shanks and St. John argue that the measures of awareness used in studies in which it is claimed that learning took place without awareness all fail to meet one or both of these criteria. For example, in studies of sequence learning, the measure of awareness may be designed to assess whether participants are aware of rules underlying the sequences, while what participants are actually aware of are permissible sequence fragments, and it is these fragments, not rules, which are responsible for participants' improved performance. In this situation, the information criterion would be unmet. On the other hand, participants may actually be aware of some kind of rule, but be unable to put it into words. In this situation, a measure which asked participants to state the rules they are aware of would fail the sensitivity criterion.

Whittlesea and Wright (1997) argue and provide evidence against separate mechanisms for implicit and explicit learning. They argue that in all learning, both implicit and explicit processes play a role. Learners are aware of the stimuli that they are processing and are aware of how they are processing them. They are, however, unaware of how their processing will influence what they learn. For example, in levels of processing experiments in which learners process stimuli either graphemically, phonemically, or semantically, learners are aware of the actual stimuli and the level at which they are processing the stimuli. They are, though, unaware that there is a higher probability that they will be able later to recognize a processed stimulus if they process it semantically rather than graphemically or phonemically. The arguments of Whittlesea and Wright are similar to arguments made by Robinson (1995b), discussed below, for different kinds of processing, not learning, in SLA

Finally, Curran and Keele (1993) reformulate the distinction between implicit and explicit learning, claiming a distinction between attentional and non-attentional learning. They compared sequential learning taking place under single-task and dual-task conditions by learners who had been informed of the sequence, by learners who had not been informed but had developed some awareness of the sequence, and by learners who had not been informed and had seemed not to have developed much awareness. It was found that while under single-task conditions more awareness coincided with more learning, under dual-task conditions differences between groups disappeared, with all groups learning very little. The researchers claim that the small amount of learning that occurred under dual-task conditions was non-attentional learning. However, as pointed out by Schmidt (1995), under the single-task condition, this study found more learning with more awareness, and under the dual-task condition the small amount of learning which took place cannot be shown to have occurred in the complete absence of awareness.

The evidence for the existence of separate mechanisms for implicit and explicit learning is not overwhelming. In addition, the distinction between implicit and explicit learning is not comparable to the distinction in SLA between acquisition and learning, at least not as the distinction is drawn by Krashen. Implicit learning is claimed to be the abstraction of structured tacit knowledge as a result attention to the surface features of a complex stimulus domain. Acquisition is claimed to take place without attention to the surface features of the second language. Also, implicit learning is seen as a general learning mechanism applicable across domains, while acquisition is seen as language specific.

Rule-based and instance-based learning. While Shanks and St. John (1994) do not find evidence of a dissociation between explicit and implicit learning, they do find evidence of a dissociation between instance-based learning and rule-based learning. For example, they argue that instance-based learning, or the memorization of strings and substrings, can explain the results of sequence learning experiments in which participants display learning but are unable to verbalize the rules underlying the sequences. In such experiments, rather than implicitly learning abstract rules, participants are able to apply their knowledge of instances. They point out, though, that there is ample evidence that rule learning is also possible, such as when participants are told something about the rules underlying the sequences.

In instance theory, Logan (1988, 1990, 1992; Logan & Etherton, 1994) explains automaticity in terms of instance-based learning. In performing non-automatic tasks, a general algorithm is used, while in automatic tasks, instances are retrieved directly from memory and applied to the task. Instances do not replace the algorithm in memory. Instead, when a task is performed, there is a race between the algorithm and the retrieval of specific instances. As the number of instances increases, so does the likelihood that one of them will win the race and the algorithm will not be applied. Logan thus views the development of automaticity as a move from rule-based to instance-based processing.

In most theories that deal with automaticity, the phenomenon will only develop after extensive practice. Logan (1990), though, argues that the same mechanism, retrieval of instances, underlies both automaticity and repetition priming. In this view, automaticity exists along a continuum. When defined as the direct retrieval and application of instances, automaticity can exist after even just a few encounters with a stimulus. The small amount of speed-up observed in studies of repetition priming is quantitatively different from the greater decrease in latency observed in studies of automaticity involving many more trials, but it is not qualitatively different in that the same process of instance retrieval underlies both phenomena.

The most recent version of adaptive control of thought theory, ACT-Rational (ACT-R) (Anderson, 1993; Anderson & Fincham, 1994; Anderson, et. al, 1997; Anderson & Labiere, 1998), gives a much stronger role to examples than had existed in ACT* (Anderson, 1983). In both ACT-R and ACT*, new knowledge always begins in a declarative form of which learners are at least potentially aware. Through the process of applying this knowledge, it can be proceduralized to produce productions which are used in skilled application of knowledge and of which learners cannot be aware. However, in ACT* declarative knowledge of instructions or rules could be used to produce productions, while in ACT-R knowledge can only be proceduralized by drawing an analogy to a declarative example. Declarative instructions and rules can only be used indirectly to access or create an example. This is one of the major changes that prompted Anderson to change the name of the theory.

A further change to the theory, although one which does not prompt a name change, appears in Anderson, et. al (1997), in which there is now a role for the direct retrieval and application of instances during performance of a cognitive skill, thus allowing for instance-based learning in ACT theory. This is a noticeable change from both ACT* and earlier versions of ACT-R, in which examples were only used for analogy before proceduralization. Skilled behavior, such as fluent use of a second language, was viewed as based completely on productions, or rules held in procedural memory. In the latest versions of ACT-R, though, skilled behavior can be either rule-based or instance-based. Proceduralized rules are extremely fast, and can become even faster as a result of strengthening through practice[2], but in situations where an exact match for a stimuli is held as an instance in declarative memory, and the instance is readily accessible, then that instance can also be applied. As in Logan, there is a race between applying rules and applying instances, and instances will be applied whenever they can win this race. It must be remembered, though, that Anderson's productions have a much better chance than Logan's algorithms of winning the race against instances and that ACT-R is still very much a rule-based system.

Complicating the distinction between rule-based learning and instance-based learning is the likelihood that the degree of similarity between novel items and previously encountered examples will affect how well learners can use examples, rather than rules, to make generalizations (e.g., Ross, 1987; Ross & Kennedy, 1990). Palmeri (1997) has tried to account for similarity by developing a model integrating Logan's instance theory of automatization and a theory of instance-based categorization (Nosofsky, 1986), called the exemplar-based random walk (EBRW) model. In this model, the degree of similarity between an instance accessed from memory and a novel item will determine the degree to which the instance helps learners categorize the novel item. The paradigm used by Palmeri (numerosity judgements) and his operationalization of similarity (distorting to a specific percent a pattern of dots used in training with a randomly generated pattern containing the same number of dots) are, of course, not applicable to studies of SLA. However, it must be recognized that similarity between novel items and previous instances may influence learners' ability to make use of instance-based rather than rule-based learning.

Neither instance-based nor rule-based learning is subliminal. Awareness is implicated in some way in both types of learning. In ACT theory (Anderson, 1983, 1993), the rules in procedural memory are not accessible to awareness, but new knowledge always begins in declarative form, which is. In instance theory (Logan, 1988; Logan & Etherton, 1994), an instance is encoded in memory if it is attended to. The amount of attention paid to an instance will determine how well it is encoded. Implicit learning (e.g., Reber, 1989; Winter & Reber, 1994) is claimed to result in tacit rule-based knowledge, but it also involves attention to the surface features of a stimulus domain. Cognitive psychology does not provide evidence for any kind of subliminal learning.

Cognitive Psychology and SLA

Researchers in SLA (e.g., Schwartz, 1986, 1993; Zobl, 1995) have tended to draw more heavily on linguistic theory, particularly UG, than on theories from cognitive psychology . This bias probably results from as well as supports the belief that language acquisition is qualitatively different from other kinds of learning. However, there also seems to be a trend in more recent research and theory toward making greater use of cognitive psychology. In the final section of this chapter, some possible connections between cognitive psychology and SLA theory are described.

As discussed above, Robinson, Schmidt, and Tomlin and Villa have all drawn on cognitive psychology in discussing the role of attention in SLA. According to Tomlin and Villa (1994), evidence from cognitive psychology indicates that detection is necessary for learning, while Schmidt (1995) expresses surprise that Tomlin and Villa were able to draw the conclusion that awareness is not necessary based on the studies that they reviewed. Schmidt (1990, 1993, 1994b, 1995) draws on cognitive psychology to argue that noticing at the level of awareness is necessary for the conversion of input to intake. Robinson (1995b) also draws on cognitive psychology to define noticing more precisely as detection with awareness as well as rehearsal in short-term memory. A few recent empirical studies, which will be discussed in the next chapter, have found empirical support for the role of noticing in SLA.

Schmidt (1992) also discusses how learning mechanisms proposed in cognitive psychology may help explain spoken second language fluency. Of the seven mechanisms reviewed by Schmidt, three, automatic and controlled processing, executive control, and restructuring, are unable to explain the power law of practice in skill development, "the most important test of any theory of cognitive learning" (p. 375). The other four are Anderson's ACT*, instance theory, strength theories, and chunking. The first of these may be used to explain how fluent L2 speakers may be able to apply proceduralized rules even though they cannot remember the declarative origins of the rules. However, Schmidt argues that there is:

> little theoretical support from psychology . . . that the development of fluency in a second language is almost exclusively a matter of the increasingly skillful application of rules. (p. 377)

The other three learning mechanisms may explain how formulas may be used in fluent speaking and how these formulas may interact with creative language production. Schmidt points out that if memory for formulas is important for fluent speech, then this would go against conventional wisdom of the peripheral role of formulaic speech.

Schmidt (1992) also points out a shortcoming of applying theories from cognitive psychology to L2 fluency:

> . . . the mechanisms made available by psychological theorizing for understanding L2 fluency derive primarily from . . . tasks that cannot be assumed to rely necessarily on the same learning mechanisms as speaking a second language. (p. 379)

However, if learning mechanisms from cognitive psychology could be used to explain SLA and second language use, then there would be no reason to claim that SLA is different from other kinds of learning.

Robinson (1995b) draws on cognitive psychology to propose two different modes of processing, both of which involve awareness, which may be used in SLA. Data-driven processing is the same as instance-based learning. It:

> . . . requires accumulation and rehearsal of instances encountered in the input in memory, and may lead to the development of simple patterns of association between co-occurring items. (pp. 301-302)

Robinson (1997b; Robinson & Ha, 1993) has also conducted studies of the role of instance-based learning in SLA. Conceptually-driven processing is a type of rule-based learning. It:

> . . . requires the elaboration of input following activation of schemata . . . and the attendant rehearsal of more abstract patterns of hierarchical organization. (p. 302)

It should be noted that this distinction is between two different kinds of processing, not between different kinds of learning. Awareness is involved in both kinds of processing. The type of processing, though, will influence what gets noticed. This is similar to the arguments of Whittlesea and Wright (1997) against separate implicit and explicit learning mechanisms.

Truscott (1998b) is somewhat unusual in that he proposes a model of SLA which draws on both Logan's instance theory, from cognitive psychology, and UG, a theory which views language acquisition as qualitatively different from other kinds of learning. In this model, UG provides the innate and invariant aspects of language while leaving "open the language-specific details, which are supplied through instance-based learning" (p. 265). Truscott then goes on to argue that Krashen's learning/acquisition distinction can be recast with learning being the development of algorithms and acquisition the accumulation of instances. Unlike Robinson (1995b), who states that instance accumulation, or data-driven processing, involves awareness, Truscott is not clear whether this accumulation of instances would involve awareness. However, since, as Robinson argues, the encoding of instances requires that they first be attended to, instance accumulation implicates awareness. In addition, since in Logan's instance theory (1988; Logan & Etherton, 1994) the application of an algorithm creates an instance which is then encoded in memory, if learning is the development of algorithms and acquisition is the accumulation of instances, there would, in Truscott's model, be an interface between learning and acquisition. Thus, Truscott's recasting of the learning/acquisition distinction does not seem compatible with the distinction as drawn by Krashen.

Conclusion

The distinction between acquisition and learning can be supported if language acquisition truly is qualitatively different from other kinds of learning and relies on a LAD which operates outside the realm of awareness. However, even if this is true, acquisition without awareness would only seem to account for the acquisition of aspects of language which fall within the scope of UG or which can be acted upon by the LAD. Other aspects of a language would have to be learned by domain-general learning mechanisms. At least some parts of each of the different domain-general learning mechanisms proposed in the field of cognitive psychology, including mechanisms for implicit learning, involve awareness. Therefore, when it comes to the acquisition of these other aspects, the acquisition/learning distinction cannot supported without recourse to subliminal learning, the existence of which is unsupported by evidence from cognitive psychology. If SLA (or even first language acquisition) is not different from other kinds of learning, as argued by Anderson (1983, 1993) and Reber (1989), and does not rely on a special LAD, then the acquisition of any aspect of language in the complete absence of awareness cannot be supported.

Chapter Two
Effects of Instruction Studies and Learner-Internal Processes

Introduction

There have been a large number of experimental studies published over the last decade looking at the effects of different kinds of instruction on SLA. While providing useful, if sometimes contradictory, information about the efficacy of different kinds of instruction, such effects of instruction studies do not always provide insight into learner-internal processes such as learner awareness or learner behavior in responding to the instruction. This chapter discusses what these studies can and cannot tell us about learner awareness and behavior, why learner awareness and behavior are important, and what needs to be done in a study in order to gain insight into these learner-internal processes.

Effects of Instruction Studies

As can be seen from Table 2.1, this decade has seen a rather large number of effects of instruction studies. This table contains information from 32 experimental studies published between 1991 and 1998, and even this is by no means comprehensive. The table is included here in order to illustrate some recent trends in the design of experimental studies of SLA.